

Metodología de análisis de relevancia de variables financieras en búsqueda de escenarios de eficiencia: Aplicación a las MIPYMES

Genaro Daza Santacoloma



Universidad Tecnológica de Pereira
Facultad de Ciencias Empresariales
Maestría en Administración Económica y Financiera
Pereira
2019

Metodología de análisis de relevancia de variables financieras en búsqueda de escenarios de eficiencia: Aplicación a las MIPYMES

Genaro Daza Santacoloma

Tesis para optar al grado de
Magíster en Administración Económica y Financiera

Director
Prof. Sandra Estrada Mejía

Universidad Tecnológica de Pereira
Facultad de Ciencias Empresariales
Maestría en Administración Económica y Financiera
Pereira
2019

Financial Feature Relevance Analysis Methodology for Searching Efficient Scenarios: Application to MSMEs

Genaro Daza Santacoloma

Thesis for the degree of
Master in economic and financial management

Advisor
Prof. Sandra Estrada Mejía

Universidad Tecnológica de Pereira
Faculty of Business Studies
Pereira
2019

Índice general

Índice general	I
Índice de tablas	III
Índice de figuras	IV
Agradecimientos	V
Resumen	VI
Abstract	VII
1. Introducción	1
2. Objetivos	5
2.1. Objetivo general	5
2.2. Objetivos específicos	5
3. Análisis de frontera eficiente	6
3.1. Concepto de eficiencia	6
3.1.1. Eficiencia técnica	7
3.1.2. Eficiencia precio	8
3.1.3. Eficiencia global	9
3.2. Análisis envolvente de datos	9
3.2.1. Orientación del modelo	10
3.2.2. Tipología de los rendimientos	10
3.3. Modelo DEA-CCR	11
3.3.1. Fijación de referencias para la mejora en DEA	13
3.4. Modelo DEA-BCC	14
3.5. Modelo DEA aditivo	15
3.6. Discusión y consideraciones acerca de DEA	16
4. Análisis de correlación canónica	18
5. Evaluación comparativa por proyección de distancia mínima	21

6. Metodología	25
6.1. Preproceso	25
6.2. Evaluación de eficiencia	27
6.3. Relevancia de variables	27
6.4. Evaluación comparativa por proyección de distancia mínima	29
7. Experimentación	30
7.1. Datos empleados	31
7.2. Resultados	33
7.2.1. Resultados de eficiencia	33
7.2.2. Resultados de relevancia de variables	36
7.2.3. Resultados de evaluación comparativa por proyección de distancia mínima	42
7.3. Discusión de los resultados y consideraciones prácticas	48
8. Conclusiones	50
9. Trabajo futuro	52
Bibliografía	53

Índice de tablas

3.1. Reemplazos duales	13
7.1. Variables de entrada y salida	30
7.2. Cantidad de empresas pequeñas por sector	31
7.3. Variables removidas en conjunto de datos	31
7.4. Dimensiones de los conjuntos de datos	31
7.5. Descripción de las variables en el sector C22.	32
7.6. Descripción de las variables en el sector C14.	32
7.7. Cantidad y porcentaje de DMUs eficientes, para C22	33
7.8. Cantidad y porcentaje de DMUs eficientes, para C14	36
7.9. Relevancia de variables para el conjunto de datos C22	40
7.10. Relevancia de variables para el conjunto de datos C14	41
7.11. Cantidad de DMUs eficientes seleccionadas como referencia	41
7.12. Empresas ineficientes para evaluación comparativa	45
7.13. Evaluación comparativa, conjunto C22	46
7.14. Evaluación comparativa, conjunto C14	47

Índice de figuras

3.1. Eficiencia técnica	7
3.2. Eficiencia precio	8
3.3. Orientación del modelo	11
3.4. Fronteras según tipología de rendimientos	14
5.1. Proyección de distancia mínima - MDM	22
6.1. Esquema metodológico	26
6.2. Distribución de probabilidad de las imágenes de CCA	28
7.1. Evaluación de eficiencia para el conjunto de datos C22	34
7.2. Evaluación de eficiencia para el conjunto de datos C14	35
7.3. Relación de entradas con eficiencia para CCR-IO, conjunto C22	36
7.4. Selección del número de variables relevantes	37
7.5. Relación de entradas con eficiencia para CCR-OO, conjunto C22	38
7.6. Relación de entradas con eficiencia para BCC-IO, conjunto C22	38
7.7. Relación de entradas con eficiencia para BCC-OO, conjunto C22	39
7.8. Relación de entradas con eficiencia para Aditivo, conjunto C22	39
7.9. Selección de las DMUs eficientes de referencia	42
7.10. Distribuciones estimadas, conjunto C22	43
7.11. DMUs eficientes seleccionadas, conjunto C22	44

Agradecimientos

Gracias a Dios por cada meta cumplida.

Deseo expresar mi gratitud a mi directora, Prof. Sandra Estrada, por su acompañamiento y motivación para realizar esta investigación. Deseo agradecer al ingeniero Jairo Castañeda por su ayuda y consejos de implementación informática. Muchas gracias a mi gran amigo el Prof. Luis Gonzalo Sánchez del colegio de ingeniería de la Universidad de Kentucky con quien pude tener una interesante discusión académica acerca de estrategias de optimización.

Resumen

En Colombia, las micro, pequeñas y medianas empresas (MIPYME) son el 96,4 % del parque empresarial nacional, generan el 80 % del empleo total y contribuyen con el 35 % del PIB. Por esta razón, es primordial avanzar en el desarrollo de herramientas, para este grupo de empresas, que les permita mejorar sus capacidades administrativas y financieras. Así, las MIPYMEs deben emplear mecanismos de evaluación y análisis que les permitan cuantificar su eficiencia, y más aún, identificar las variables e interacciones, que contribuyen en el incremento la eficiencia.

En condiciones prácticas, es un problema vigente establecer los objetivos de intervención, sobre las variables financieras, que les permitan a las empresas conseguir mejores resultados en sus procesos. En tal sentido, en este trabajo se presenta el desarrollo de una metodología de análisis de relevancia de variables financieras por medio de la combinación de técnicas de análisis de frontera eficiente con técnicas de aprendizaje de máquina.

La metodología desarrollada en este documento se soporta en tres fases principales. La primera fase consiste en el cálculo de la eficiencia relativa de un conjunto de empresas, representadas a través de variables de entrada y salida, empleando análisis envolvente de datos (DEA). En la segunda fase se determina cuáles de las variables de entrada tienen mayor relación con la eficiencia calculada, estableciendo un conjunto reducido de variables relevantes, lo cual se realiza mediante el análisis de correlación canónica (CCA). Finalmente, en la tercera fase, proponemos un novedoso método de evaluación comparativa, denominado proyección de distancia mínima (MDM), el cual posibilita calcular los valores de cambio para cada una de las variables (relevantes) de las empresas ineficientes, de manera que éstas puedan asemejarse al conjunto de empresas eficientes.

Abstract

In Colombia, the micro, small, and medium enterprises (MSME) are the 96,4% of the whole national firms, they produce the 80 % of the total work and the 35 % of the GDP. That is why, it is primordial to develop tools for improving administrative and financial capabilities of the MSME. In this sense, the MSMEs should use assessment mechanisms for quantifying their efficiency and, moreover, identifying the variables and interactions that helps to the increase of the efficiency.

Under practical constraints, it is still a problem to determine the intervention targets for the financial variables that allow to outperform the enterprise process. Then, in this work is developed a financial feature relevance analysis methodology by means of combining efficient frontier analysis and machine learning techniques.

The proposed methodology is based on three main stages. The first stage consists in computing the relative efficiency for a set of firms represented by both input and output variables. The efficiency is computed using Data Envelopment Analysis (DEA). In the second stage, the variables having a greater relationship with the computed efficiency are established, which produces a reduced set of relevant features. This procedure is carry out by means of Canonical Correlation Analysis (CCA). Finally, in the third stage, we propose a novel benchmarking method, that we named Minimum Distance Mapping (MDM), which allow to calculate, for the inefficient firms, the amount of change that each one of the relevant variables requires, in order to achieve to be as the efficient firms.

Capítulo 1

Introducción

En Colombia, las micro, pequeñas y medianas empresas (MIPYME) son el 96,4 % del parque empresarial nacional (92.6 % microempresas y 3.8 % son pequeñas y medianas empresas), las MIPYMEs generan el 80 % del empleo total y contribuyen con el 35 % del PIB [Espinosa et al.2015, Dinero2017]. Es fundamental brindar, a este grupo de empresas, capacidades administrativas y sobre todo financieras, para que su trabajo (producción o servicios) se realice de manera adecuada y logren generar valor.

Como es sabido, las empresas más exitosas están permanentemente mejorando sus productos y servicios a través de procesos de innovación, pero también a través de la revisión constante de sus modelos administrativos y financieros, permitiéndoles aprovechar al máximo los recursos disponibles. Así pues, las empresas exitosas están en búsqueda de la eficiencia.

Adicionalmente, el objetivo fundamental de las empresas de lucro es generar constantemente más recursos financieros, por lo cual la evaluación del cumplimiento de ese objetivo es un elemento esencial de la gestión financiera y administrativa. Las MIPYMEs deben analizar sus condiciones y entorno financiero para no tener dificultades por falta de capacidad para la obtención de recursos financieros, es decir, las MIPYMEs deben emplear mecanismos de evaluación que les permitan cuantificar la eficiencia, y más aún, identificar las variables y los mecanismos que contribuyen en el incremento la misma.

A partir del registro de los estados financieros se puede construir un conjunto de indicadores que ayudan a interpretar de forma concisa la situación financiera de la empresa. Estos indicadores son principalmente de cuatro tipos: apalancamiento, liquidez, rentabilidad y eficiencia. En principio, los indicadores financieros pueden dar una idea de las situaciones que ocurren al interior de la empresa, de su estado de endeudamiento, de su capacidad de producción, de su posibilidad de generar recursos, etc., sin embargo, el análisis no articulado de tales indicadores puede llevar a la toma de decisiones parcializada e inclusive contradictoria. Así, el problema al trabajar con los indicadores es que estos observan la empresa por partes y no la consideran un todo [Feroz et al.2003, Fenyves et al.2015, Hui et al.2015]. Asimismo, en [Jakub et al.2015] se señala que indicadores como MVA y EVA han sido preponderantes en las empresas para la evaluación de la efectividad de las actividades del negocio.

Con el fin de interpretar de una manera integral los resultados de la empresa, desde hace ya varias décadas, se han propuesto indicadores generalizantes para evaluar el valor

o la capacidad de generación de valor de la empresa, tales como: el beneficio antes de intereses e impuestos (EBIT - *Earnings Before Interest and Taxes*), el beneficio antes de intereses, impuestos, depreciación y amortización (EBITDA - *Earnings before Interest, Taxes, Depreciation and Amortization*) [Méndez2007], valor de mercado agregado (MVA - *Market Value Added*) [Lee and Kim2009], el valor económico agregado (EVA - *Economic Value Added*) [O'Hanlon and Peasnell1998], entre otros. Por otra parte, también se han formulado metodologías para determinar la eficiencia de la actividad empresarial, tales como: análisis envolvente de datos (DEA - *Data Envelopment Analysis*) [Ray2004], y el análisis estocástico de frontera (SFA - *Stochastic Frontier Analysis*) [Khalil et al.2015].

En tal sentido, varias investigaciones han comparado los resultados de las metodologías basadas en beneficios (EBIT y EBITDA) contra las metodologías de valor agregado (MVA y EVA). Así, en [Biddle et al.1997] se asegura que los componentes del EVA se agregan sólo marginalmente a la información contenida más allá de los beneficios, y que los indicadores basados en beneficios están más altamente asociados con los retornos y con el valor de la compañía que el resultado de EVA. Por el contrario, en investigaciones más recientes, algunos autores defienden la postura que teóricamente el indicador EVA es superior a los indicadores convencionales (EBIT y EBITDA) debido a que éste considera el costo total del capital, incluyendo el costo del patrimonio, además es una medida de la utilidad económica pura [de Wet2005, Shil2009, Saha et al.2016, Feltham et al.2004].

Por otra parte, tanto EVA como DEA han sido empleados para evaluar la eficiencia del capital, y en [Shi2012] se señala que DEA refleja de mejor manera la eficiencia del capital, y que los resultados que arroja EVA pueden ser engañosos y llevar a decisiones de inversión inadecuadas. Esta frase es también soportada por [Hui et al.2015] donde se expresa que el EVA de las compañías puede estar subestimado llevando a decisiones de inversión inapropiadas. En concordancia con estas opiniones, en este trabajo la evaluación del desempeño se desarrolla a través de técnicas de análisis de frontera eficiente, que reflejan el potencial de la empresa para maximizar las utilidades, dadas una entradas determinadas.

Con relación a DEA, se puede decir que es la técnica de análisis de eficiencia más empleada, debido a que su implementación es sencilla, la interpretación de sus resultados es natural, y además porque es un modelo no paramétrico. En la literatura se ha planteado que uno de sus inconvenientes radica en que esta técnica no establece medidas, asociaciones estadísticas o relaciones de causalidad entre las variables de entrada y salida [Sexton et al.1986]. Sin embargo, tal afirmación puede ser parcialmente cierta, dado que se conoce un esquema de trabajo en el que se emplea DEA como estrategia, no solo para identificar las variables más significativas en el logro de la eficiencia de cada empresa, sino también para establecer los valores ideales de tales variables, de forma que la empresa pueda ser eficiente [Serrano and Blasco2000], a esto se le conoce como evaluación comparativa o *benchmarking*-DEA. La dificultad con el uso de *benchmarking*-DEA es que la identificación de las variables relevantes (y sus valores de cambio) sólo puede hacerse bajo la hipótesis que la empresa logra la eficiencia técnica relativa máxima, lo cual, en condiciones prácticas, cotidianas y reales puede ser inalcanzable.

Adicionalmente, también se han formulado alternativas basadas en análisis de correlación y regresión sobre las variables de entrada [Nataraja and Johnson2011, Shi2012], y la articulación con métodos heurísticos de búsqueda secuencial por agregación o remoción

de variables de entrada [Wagner and Shimshak2007]; estas alternativas permiten establecer la pertinencia de las variables de entrada para que sean consideradas durante el análisis y así mejorar la eficiencia.

En síntesis, la literatura señala argumentos contradictorios (en favor y en contra) de cada una de las propuestas metodológicas de evaluación de desempeño financiero o de eficiencia del capital, y adicionalmente, en las propuestas metodológicas de evaluación de eficiencia financiera, encontradas en el estado del arte, es un problema vigente la identificación de las variables de entrada y la medida en que éstas deben ser intervenidas, en condiciones prácticas, para obtener mejores resultados sobre las variables de salida, mejorando de esta forma el nivel de eficiencia de la empresa.

Por todas las razones expuestas, en este trabajo se presenta una metodología que permite identificar cuáles variables de entrada, y en qué proporción estas variables, deben ajustarse para lograr un efecto positivo sobre la eficiencia, aplicado a casos prácticos de empresas PYMEs.

La metodología desarrollada en este documento se soporta en tres fases principales:

1. **Evaluación de eficiencia:** A partir de un grupo de empresas de condiciones semejantes, es decir, de igual sector económico y con cantidad similar de activos totales (tamaño empresarial), para las cuales se conoce un conjunto de variables financieras de entrada y salida, se realiza el cálculo de la eficiencia empleando análisis envolvente de datos (DEA), de manera que sea posible clasificar las empresas en eficientes e ineficientes.
2. **Relevancia de variables:** Con las variables de entrada y las etiquetas de eficiencia, se realiza análisis de correlación canónica (CCA - *Canonical Correlation Analysis*) para determinar cuáles de las variables de entrada originales tienen mayor relación con la eficiencia de las empresas. Adicionalmente, CCA permite establecer un conglomerado de empresas eficientes y/o cuasi-eficientes, que sirvan como base de referencia y ejemplo para las empresas ineficientes.
3. **Evaluación comparativa:** Busca establecer los valores de cambio para cada una de las variables relevantes de una empresa ineficiente, de manera que la empresa ineficiente pueda asemejarse al conglomerado de empresas eficientes y/o cuasi-eficientes. Para lograr este propósito, en este trabajo y como factor de novedad, se propone el método de proyección de distancia mínima (MDM - *Minimum Distance Mapping*), el cual consiste en proyectar en el espacio de variables relevantes la empresa ineficiente sobre la envolvente convexa determinada por el conglomerado de empresas eficientes y/o cuasi-eficientes.

La estructura de este documento es la siguiente: El capítulo 2 presenta el objetivo general y los objetivos específicos de esta investigación. El capítulo 3 presenta los fundamentos teóricos del análisis de frontera eficiente, mayoritariamente tomados del libro [Serrano and Blasco2000], explicando los conceptos de eficiencia y el cálculo del análisis envolvente de datos (DEA). El capítulo 4 brinda la base teórica del análisis de correlación canónica (CCA), el cual es una recopilación y simplificación del artículo [Uurtio et al.2018]. En el capítulo 5 proponemos y formulamos la estrategia de evaluación comparativa para las

empresas ineficientes por medio del método que hemos denominado proyección de distancia mínima (MDM). El capítulo 7 contiene el marco experimental, donde se presentan y discuten los resultados del método propuesto sobre bases de datos de un conjunto de empresas. El capítulo 8 exhibe las conclusiones del trabajo y finalmente el capítulo 9 presenta el trabajo futuro que se puede desprender de esta investigación.

Capítulo 2

Objetivos

2.1. Objetivo general

Desarrollar una metodología de análisis de relevancia de variables financieras por medio de la combinación de técnicas de análisis de frontera eficiente con técnicas de aprendizaje de máquina para la proyección de observaciones en espacios de dimensión reducida, que permita establecer las variables que se deben intervenir en pro de optimizar la eficiencia de las empresas.

2.2. Objetivos específicos

- Calcular la eficiencia de un conjunto de empresas que pertenecen a un mismo sector de la economía aplicando técnicas de análisis de frontera eficiente que permita categorizar o etiquetar las empresas (observaciones) evaluadas según su nivel de eficiencia.
- Determinar, empleando técnicas de aprendizaje de máquina, una proyección de las observaciones de un espacio de características (variables de entrada) a un espacio de dimensión reducida en el cual las observaciones se agrupan según su etiqueta de eficiencia.
- Establecer las variables más relevantes, que se deben intervenir en pro de optimizar la eficiencia de la empresa, empleando una función de traslación de una observación del agrupamiento de menor eficiencia al agrupamiento de mayor eficiencia dentro del espacio de dimensión reducida, sujeta a las restricciones de las variables originales.

Capítulo 3

Análisis de frontera eficiente

El análisis de frontera es una medida cuantitativa, determinada objetivamente, que remueve los efectos de los precios de mercado y otros factores exógenos que influyen en el desempeño observado; este análisis permite la incorporación de múltiples entradas y salidas para medir el desempeño y proveer un valor de referencia, i.e. la frontera, contra la cual los competidores (empresas) pueden identificar regiones de mejores y peores prácticas asociadas con la medida de desempeño. Así, el análisis de frontera permite a los administradores identificar la brecha entre su desempeño actual y el desempeño óptimo [Assaf and Josiassen2016].

3.1. Concepto de eficiencia

La eficiencia es un concepto bastante usado en la economía. En general, el concepto básico de eficiencia en economía se refiere a la relación entre entrada y salida o costos y beneficios [Shi2012]. Cuando el concepto de eficiencia es aplicado a una empresa, señala si la empresa obtiene el máximo rendimiento de sus recursos, o si una cierta cantidad de rendimiento se logra a un costo mínimo.

La eficiencia es la capacidad para realizar o cumplir adecuadamente una función, es decir, cumplir un objetivo optimizando los recursos disponibles. Cuando se habla de la eficiencia de una empresa, usualmente nos referimos al éxito en producir tanto como sea posible de una salida a partir de un conjunto de entradas dadas [Farrell1957]. Sin embargo, la eficiencia también puede lograrse al disminuir tanto como sea posible las entradas, sin sacrificar la cantidad de producción de la salida.

La medición de la eficiencia permite entonces proporcionar una base objetiva para evaluar el desempeño y la administración. El resultado en el nivel más alto de eficiencia (por ejemplo, el máximo beneficio posible) proporciona un estándar absoluto para la gestión por objetivos [Ray2004].

Para establecer la eficiencia de las empresa son generalmente tenidos en cuenta 3 supuestos básicos [Serrano and Blasco2000]:

1. Se conoce la función de producción eficiente.
2. Las empresas operan bajo condiciones de rendimientos constantes a escala, es decir, el incremento porcentual de la salida es igual al incremento porcentual de las entradas.

3. Mayores incrementos en las entradas por unidad de salida corresponden con eficiencias técnicas más baja.

Sin embargo, estos supuestos pueden relajarse considerando rendimientos crecientes y decrecientes a escala.

Por otra parte, existen diferentes puntos de vista de la eficiencia y diferentes descripciones teóricas; en tal sentido, podemos encontrar: la eficiencia técnica, la eficiencia precio y la eficiencia global.

3.1.1. Eficiencia técnica

La eficiencia técnica refleja la capacidad de una empresa para obtener la máxima salida a partir de un conjunto dado de entradas.

Suponga que se tienen 4 unidades de toma de decisión (DMU – *Decision Making Unit*) o empresas, que se denominarán A, B, C y D. Cada DMU tiene una única salida y que requiere de dos entradas x_1 y x_2 . Considerando la Figura 3.1, en la cual cada punto representa las coordenadas del plan de producción observado para cada una de las DMU. La curva isocuanta dada por II' corresponde a las DMU eficientes, tal que las DMU que estén por encima de esta curva resultan ser ineficientes. Se puede ver que las DMUs B y D son ineficientes técnicamente. Ambas DMU podrían reducir la cantidad de entradas consumidas y seguir produciendo una unidad de salida. La ineficiencia de estas DMU está dada por la distancia $B'B$ y $D'D$, respectivamente. Por otra parte, las DMUs A y C son técnicamente eficientes puesto que operan sobre la curva isocuanta eficiente.

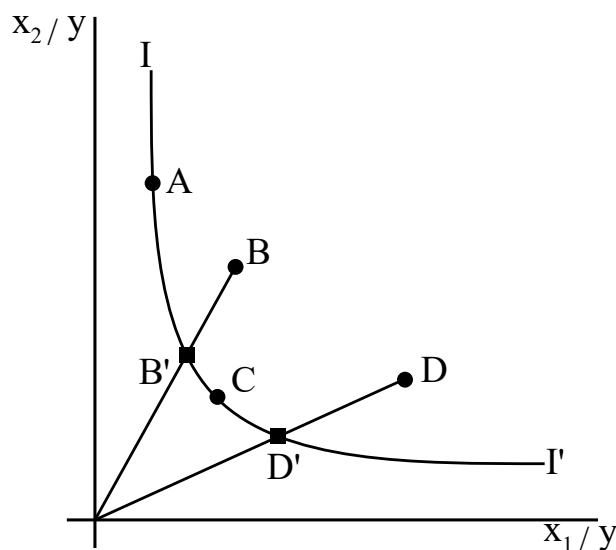


Figura 3.1: Eficiencia técnica

Dado lo anterior, puede obtenerse el valor de eficiencia relativa como la relación entre la longitud de la línea que va desde el origen hasta el punto de intersección y la longitud de la línea que une el origen con la DMU considerada. El punto de intersección corresponde al punto de cruce entre la frontera eficiente y la recta que une el origen con la DMU

respectiva. Para la empresa B si tiene:

$$ET_B = \frac{\text{dist}(O, B')}{\text{dist}(O, B)} \quad (3.1)$$

donde ET_B es la eficiencia técnica de B y $\text{dist}(\cdot, \cdot)$ es la distancia entre los puntos.

De la ecuación (3.1), la eficiencia técnica sólo puede tomar valores comprendidos entre cero y uno. Un valor cercano a cero significa que la DMU evaluada está muy lejos de la curva isocuanta eficiente y por tanto se trata de una DMU técnicamente ineficiente. Lo contrario ocurre cuando la eficiencia técnica está cerca a uno, caso de A y C.

3.1.2. Eficiencia precio

La eficiencia de precio se refiere a la capacidad de la DMU para usar las distintas entradas en proporciones óptimas dados sus precios relativos. En la Figura 3.2 se presenta la línea PP' , llamada línea isocoste; la pendiente de esta línea presenta la relación entre los precios de las entradas x_1 y x_2 .

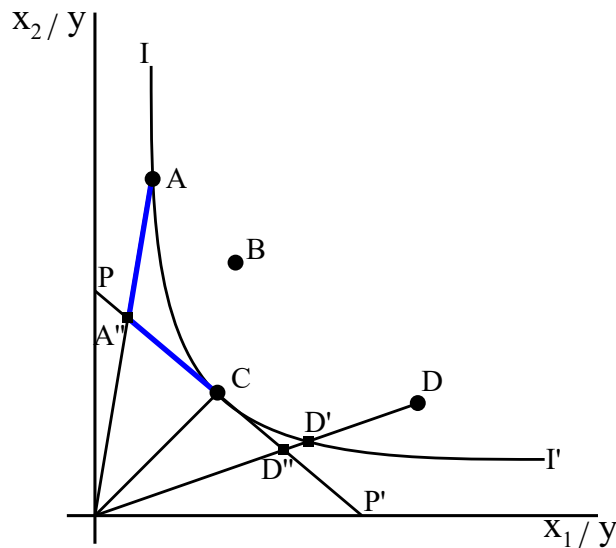


Figura 3.2: Eficiencia precio

Como se mencionó antes, las DMUs A y C presentan eficiencia técnica porque operan sobre la curva isocuanta. Sin embargo, como puede verse en la Figura 3.2, sólo la DMU C resulta ser también eficiente en precios, mientras que la DMU A debe reducir los precios totales en la distancia $A''A$, o alternatively en la proporción

$$\left(1 - \frac{\text{dist}(O, A'')}{\text{dist}(O, A)}\right) \cdot 100.$$

El valor de eficiencia precio se obtiene como la relación entre la longitud de la línea que va desde el origen hasta el punto proyectado sobre la curva isocoste y la longitud de

la línea que va desde el origen hasta el punto proyectado sobre la curva isocuanta. Por ejemplo para la DMU A se tiene,

$$EP_A = \frac{\text{dist}(O, A'')}{\text{dist}(O, A)}, \quad (3.2)$$

donde EP_A es la eficiencia precio de A.

EP puede tomar valores comprendidos entre cero y uno, y mientras más lejos se esté de uno se considerará que la DMU es más ineficiente en precios.

3.1.3. Eficiencia global

La eficiencia global, también llamada eficiencia económica, se obtiene mediante el cociente entre la longitud de la línea que va desde el origen hasta el punto proyectado sobre la línea isocoste eficiente y la longitud de la línea que va desde el origen hasta el punto que representa la DMU considerada. Por ejemplo, considere la DMU D de la Figura 3.2, la eficiencia global está dada por:

$$EG_D = \frac{\text{dist}(O, D'')}{\text{dist}(O, D)}. \quad (3.3)$$

La eficiencia global, expresada en la ecuación (3.3), se puede descomponer como:

$$EG_D = \frac{\text{dist}(O, D')}{\text{dist}(O, D)} \cdot \frac{\text{dist}(O, D'')}{\text{dist}(O, D')}. \quad (3.4)$$

La ecuación (3.4) significa que la eficiencia global es igual al producto de la eficiencia técnica y la eficiencia precio, y su valor está comprendido entre cero y uno.

En este trabajo en particular, las eficiencias calculadas serán eficiencias técnicas, debido a que el cálculo de la eficiencia precio y por ende de la eficiencia global requiere de un conocimiento a priori de los precios (ponderación) de las entradas y las salidas, que para los casos a analizar no están disponibles.

3.2. Análisis envolvente de datos

El análisis envolvente de datos (DEA) es un método no paramétrico para medir la eficiencia de unidades de toma de decisión (DMU), tales como una empresa o una agencia estatal. En un principio los modelos DEA fueron empleados para evaluar la eficiencia relativa de organizaciones sin ánimo de lucro, pero su practicidad, popularidad y versatilidad, permitió que su uso se extendiera rápidamente al análisis en organizaciones lucrativas.

DEA acepta el uso de múltiples entradas y salidas de forma simultánea, sin ninguna consideración acerca de la distribución de los datos, y la eficiencia es medida en términos de un cambio proporcional en las entradas o salidas. La medición de eficiencia se realiza mediante el uso de técnicas de programación lineal consistentes en envolver los vectores observados de entrada salida lo más estrechamente posible [Ray2004, Ji and Lee2010].

A pesar de que en el trabajo de [Farrell1957] se supone que la frontera de producción es conocida, en la práctica esto no es así y por tanto es necesario estimarla. El algoritmo

de DEA permite la construcción de una superficie envolvente, frontera eficiente o función de producción empírica, a partir de los datos disponibles del conjunto de DMUs objeto de estudio, de forma que las DMUs que determinan la envolvente son denominadas unidades eficientes y aquellas que no permanecen sobre la envolvente son consideradas unidades ineficientes [Serrano and Blasco2000].

La medida de la eficiencia de una DMU mediante la técnica DEA implica dos pasos básicos:

1. Construcción del conjunto de posibilidades de producción.
2. Estimación de la máxima expansión factible de las salidas o de la máxima contracción de las entradas de la DMU del conjunto de posibilidades de producción.

Los modelos DEA se pueden clasificar en función del tipo de medida de eficiencia que proporcionan (radiales y no radiales), la orientación del modelo (orientado a entradas y orientado a salidas), y la tipología de los rendimientos a escala. En este documento particularmente solo se considerarán modelos DEA radiales.

3.2.1. Orientación del modelo

Orientado a entradas: busca la máxima reducción proporcional en el vector de entradas, satisfaciendo al menos el nivel de salidas dado.

Orientado a salidas: busca el máximo incremento proporcional de las salidas sin requerir más valor de entradas que las observadas.

Teniendo en cuenta las orientaciones, una DMU será considerada eficiente si, y sólo si, no es posible incrementar las cantidades de salida manteniendo fijas las cantidades de entrada utilizadas, ni es posible disminuir las cantidades de entrada empleadas sin alterar las cantidades de salida obtenidas.

La Figura 3.3 presenta el caso de una única entrada y una única salida, bajo la suposición de rendimientos constantes a escala. La DMU A es ineficiente técnicamente puesto que está situada por debajo de la frontera (línea diagonal).

Para un modelo orientado a entradas, la DMU A, puede reducir la cantidad de entrada x y seguir produciendo la misma cantidad de salida y , tomando como referencia la mejor práctica de la DMU A_1 y la eficiencia técnica sería $ET_A = BA_1/BA$. Para el modelo orientado a salidas, la DMU A, es ineficiente, porque puede producir mayor cantidad de salida consumiendo la misma cantidad de entrada, en este caso la eficiencia es $ET_A = CA/CA_2$. Para el supuesto de rendimientos constantes a escala, las medidas de eficiencia técnica de entradas y salidas son iguales.

3.2.2. Tipología de los rendimientos

Los rendimientos a escala pueden ser constantes, crecientes o decrecientes e indican los incrementos de la producción que son resultado de los incrementos de todos los factores de producción en el mismo porcentaje.

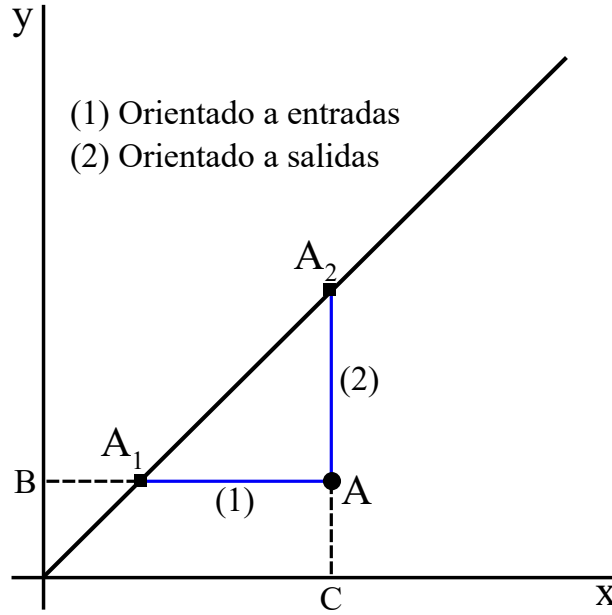


Figura 3.3: Orientación del modelo

Rendimientos constantes a escala: cuando el incremento porcentual de la salida es igual al incremento porcentual de los recursos productivos.

Rendimientos crecientes a escala: también llamados economías de escala, se presentan cuando el incremento porcentual de la salida es mayor que el incremento porcentual de los recursos productivos.

Rendimientos decrecientes a escala: cuando el incremento porcentual de la salida es menor que el incremento de los recursos productivos.

3.3. Modelo DEA-CCR

Su nombre se debe a que fue desarrollado por Charnes, Cooper y Rhodes [Charnes et al.1978]. El modelo DEA-CCR proporciona medidas de eficiencia radiales, es orientado a entradas o a salidas, supone convexidad y rendimientos constantes a escala.

La eficiencia técnica relativa de cada DMU se define como el cociente entre la suma ponderada de las salidas y la suma ponderada de las entradas, sujeto a que ninguna DMU puede tener una eficiencia mayor que uno, de manera que para el caso de un modelo orientado a entradas se tiene:

$$\max_{u,v} h_0 = \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}} \quad \text{sujeto a :} \quad \begin{cases} \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1 & j = 1, 2, \dots, n \\ u_r, v_i \geq 0 \end{cases} \quad (3.5)$$

donde se consideran n DMUs ($j = 1, 2, \dots, n$), cada una de las cuales tiene las mismas entradas (en diferentes cantidades) para obtener las mismas salidas (en diferentes

cantidades). La cantidad de la i -ésima entrada ($i = 1, 2, \dots, m$) consumidas por la j -ésima DMU se representan por x_{ij} con ($x_{ij} \geq 0$). La cantidad de la i -ésima entrada consumida por la DMU que se está evaluando es x_{i0} . La cantidad observada de la r -ésima salida ($r = 1, 2, \dots, s$) producida por la j -ésima DMU se representa por y_{rj} con ($y_{rj} \geq 0$). La cantidad de la r -ésima salida producida por la DMU que se está evaluando es y_{r0} . Finalmente, u_r para ($r = 1, 2, \dots, s$) y v_i para ($i = 1, 2, \dots, m$) representan los pesos de ponderación de las salidas y las entradas, respectivamente.

La ecuación (3.5) es un problema no lineal que busca obtener el conjunto óptimo de pesos (u_r y v_i) que maximicen la eficiencia relativa h_0 de la DMU en evaluación. Si $h_0 = 1$, la DMU evaluada es eficiente, si $h_0 < 1$, la DMU evaluada es ineficiente. En este caso las DMUs con los mismos pesos u_r y v_i asignados a la DMU ineficiente que está siendo evaluada que resulten ser eficientes se denominan pares. Los pares constituyen el conjunto de referencia de la DMU ineficiente, es decir, la referencia para la mejora de la DMU ineficiente.

La ecuación (3.5) puede modificarse sustituyendo la condición de no negatividad $u_r, v_i \geq 0$ por una condición de positividad estricta $u_r, v_i \geq \varepsilon$, donde ε es un infinitésimo no-arquimedeo, la razón de este cambio es evitar que una DMU, pese a presentar $h_0 = 1$, sea incorrectamente caracterizada como eficiente al obtener una solución en la que algún peso u_r y/o v_i sea cero, y por tanto la respectiva entrada o salida sea obviada al determinar la eficiencia. Así se tiene que:

$$\max_{u,v} h_0 = \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}} \quad \text{sujeto a :} \quad \begin{cases} \frac{\sum_{r=1}^s u_r y_{rj}}{m} \leq 1 & j = 1, 2, \dots, n \\ u_r, v_i \geq \varepsilon \end{cases} \quad (3.6)$$

Un inconveniente de la ecuación (3.6) es que puede tener un número infinito de soluciones, puesto que si u_r^*, v_i^* es una solución $h_0 = 1$, entonces $\beta u_r^*, \beta v_i^*$ también lo es para todo $\beta > 0$.

Este problema puede linealizarse y expresarse en forma matricial para llegar a la siguiente expresión [Serrano and Blasco2000]:

$$\max_{\mu, \delta} w_0 = \mu^\top \mathbf{y}_0 \quad \text{sujeto a} \quad \begin{cases} \delta^\top \mathbf{x}_0 = 1 \\ \mu^\top \mathbf{Y} - \delta^\top \mathbf{X} \leq 0 \\ \mu^\top, \delta^\top \geq \varepsilon \end{cases} \quad (3.7)$$

donde \top representa la transpuesta, ε es un vector con elementos iguales positivos de tipo infinitésimo no-arquimedeo, $\mathbf{Y} \in \mathbb{R}^{s \times n}$ es una matriz de salidas, \mathbf{y}_0 representa el vector de salidas de la DMU que se está evaluando, $\mathbf{X} \in \mathbb{R}^{m \times n}$ es la matriz de entradas, \mathbf{x}_0 representa el vector de entradas de la DMU que se está evaluando, $\mu \in \mathbb{R}^{s \times 1}$ es el vector de pesos de la salidas y $\delta \in \mathbb{R}^{m \times 1}$ es el vector de pesos asociado a las entradas.

En este caso, la DMU se califica eficiente si $w_0 = 1$ y los pesos $\mu, \delta > 0$. Si la DMU es ineficiente se tiene que $w_0 < 1$ con pesos μ^*, δ^* y existirá al menos una DMU que satisface la restricción

$$\sum_{r=1}^s \mu_r y_{rj} = \sum_{i=1}^m \delta_i x_{ij},$$

para los mismos pesos μ^*, δ^* . El conjunto de las DMUs que satisfacen la restricción son eficientes y constituyen el conjunto referencia de la DMU evaluada, siendo la existencia de estas DMUs eficientes la que fuerza a la DMU evaluada a ser ineficiente [Cooper et al.2000].

Existe una tercera representación para el modelo, conocida como forma envolvente, la cual se obtiene al emplear los reemplazos duales para las condiciones primales del modelo (3.7), tales reemplazos se presentan en la tabla la Tabla 3.1.

Restricción primal	Variable dual	Restricción dual	Variable primal
$\delta^\top \mathbf{x}_0 = 1$	Θ	$\mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{y}_0$	$\delta^\top \geq \mathbf{0}$
$\mu^\top \mathbf{Y} - \delta^\top \mathbf{X} \leq \mathbf{0}$	$\boldsymbol{\lambda} \geq \mathbf{0}$	$\Theta \mathbf{x}_0 - \mathbf{X}\boldsymbol{\lambda} \geq \mathbf{0}$	$\mu^\top \geq \mathbf{0}$

Tabla 3.1: Reemplazos duales

Empleando los reemplazos se obtiene:

$$\min_{\Theta, \boldsymbol{\lambda}} z_0 = \Theta \quad \text{sujeto a} \quad \begin{cases} \mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{y}_0 \\ \Theta \mathbf{x}_0 \geq \mathbf{X}\boldsymbol{\lambda} \\ \boldsymbol{\lambda} \geq \mathbf{0} \end{cases} \quad (3.8)$$

donde $\boldsymbol{\lambda} \in \mathbb{R}^{n \times 1}$ es el vector de pesos y λ_j es el peso de la j -ésima DMU, Θ denota la eficiencia técnica de la DMU evaluada.

Con el modelo DEA puede obtenerse para toda DMU ineficiente un punto de proyección sobre la frontera eficiente que represente a una DMU (real o virtual) que sea eficiente, lo que para un modelo orientado a entradas implica que consumirá a lo sumo una proporción de las entradas originales y producirá la misma cantidad de salida. Las DMUs implicadas en la construcción de la DMU eficiente referida (real o virtual) constituyen el conjunto de referencia de la DMU evaluada y calificada como ineficiente. El conjunto de referencia estará formado por todas las DMUs que en la solución de (3.8) obtengan pesos $\lambda_j^* > 0$ con $j = 1, 2, \dots, n$.

En general las ecuaciones (3.6), (3.7) y (3.8), que describen un modelo DEA-CCR orientado a las entradas, pueden reformularse como modelos orientados a salidas con expresiones de complejidad semejante, ver detalles en [Serrano and Blasco2000].

3.3.1. Fijación de referencias para la mejora en DEA

De la ecuación (3.8), el punto de proyección sobre la frontera eficiente, es una combinación lineal de los puntos observados y está dado por $\hat{\mathbf{x}}_0 = \boldsymbol{\lambda}^* \mathbf{X}$, $\hat{\mathbf{y}}_0 = \boldsymbol{\lambda}^* \mathbf{Y}$.

Las coordenadas de la proyección sobre la frontera eficiente representan los valores de entrada y salida objetivo para la DMU ineficiente, es decir, los niveles de entrada y salida que la convertirán en eficiente en caso de alcanzarlos, y comparando los valores de la DMU evaluada con los valores objetivo es posible establecer la cantidad (absoluta y/o relativa) de la reducción de las entradas y/o incrementos de las salidas necesarios para convertirse en eficiente.

3.4. Modelo DEA-BCC

El nombre del modelo DEA-BCC se debe a que sus desarrolladores son Banker, Charnes y Cooper [Banker et al.1984]. A diferencia del modelo DEA-CCR que considera rendimientos constantes a escala, el modelo DEA-BCC flexibiliza el análisis, permitiendo que la tipología de rendimiento a escala en un momento dado sea constante, creciente o decreciente.

Los rendimientos a escala indican como los incrementos de producción son resultado del incremento de todos los factores de producción en el mismo porcentaje (ver Sección 3.2.2).

En la Figura 3.4 se presentan las fronteras estimadas para rendimientos variables a escala (RVE), rendimientos no crecientes a escala (RNCE) y rendimientos constantes a escala (RCE).

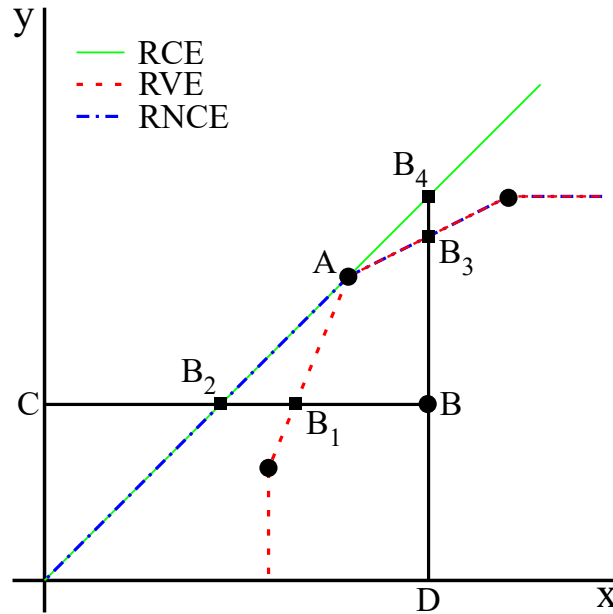


Figura 3.4: Fronteras según tipología de rendimientos

Considerando un modelo orientado a entradas bajo el supuesto de rendimientos constantes o no crecientes a escala, puede verse que la eficiencia de la DMU B es el coeficiente $\text{dist}(C, B_2)/\text{dist}(C, B)$, para el caso de rendimientos variables la eficiencia es $\text{dist}(C, B_1)/\text{dist}(C, B)$. La diferencia entre ambas medidas de eficiencia, es decir la distancia B_2B_1 , corresponde con la eficiencia escala (EE), y se interpreta como la parte de la ineficiencia presente en la eficiencia técnica global (ETG) que obedece a la escala de producción de la DMU, tal que:

$$\begin{aligned} \text{ETG} &= (\text{ETP})(\text{EE}) \\ \frac{\text{dist}(C, B_2)}{\text{dist}(C, B)} &= \frac{\text{dist}(C, B_1)}{\text{dist}(C, B)} \cdot \frac{\text{dist}(C, B_2)}{\text{dist}(C, B_1)}, \end{aligned} \quad (3.9)$$

donde ETP es la eficiencia técnica pura y puede inferirse que si $\text{EE} = 1$ entonces $\text{ETG} = \text{ETP}$, indicando que la DMU no presenta ineficiencias de escala. Cuando la DMU presenta

ineficiencias de escala ($EE < 1$) se debe comparar la frontera de rendimientos variables con la frontera de rendimientos no crecientes para determinar si la DMU opera bajo rendimientos crecientes o decrecientes a escala.

Por otra parte, la frontera RCE produce generalmente un número menor de DMUs eficientes y puntuaciones de eficiencia menores. Además, para la frontera RVE la eficiencia orientada a entradas no es necesariamente igual que la eficiencia orientada a salidas.

Similarmente que en el modelo DEA-CCR, para el modelo DEA-BCC, orientado a entradas, se tiene el siguiente modelo de optimización:

$$\max_{\mathbf{u}, \mathbf{v}, \mathbf{k}} h_0 = \frac{\mathbf{u}^\top \mathbf{y}_0 + \mathbf{k}_0}{\mathbf{v}^\top \mathbf{x}_0} \quad \text{sujeto a} \quad \begin{cases} \frac{\mathbf{u}^\top \mathbf{Y}_j + \mathbf{k}_0}{\mathbf{v}^\top \mathbf{x}_0} \leq 1 & j = 1, 2, \dots, n \\ \mathbf{u}^\top, \mathbf{v}^\top \geq \mathbf{I}_\varepsilon \\ \mathbf{k}_0 \text{ no restringida} \end{cases} \quad (3.10)$$

La diferencia entre la ecuación (3.10) y la del modelo DEA-CCR (3.6) radica en que para DEA-BCC el numerador de las salidas ponderadas está adicionado con un término constante \mathbf{k}_0 , que en el caso DEA-CCR valdría cero y representa los rendimientos constantes a escala, de modo que la solución óptima de (3.10) para la DMU que se supone eficiente conlleva a:

$\mathbf{k}_0^* > 0$ para todas las soluciones óptimas, prevalecen rendimientos crecientes a escala.

$\mathbf{k}_0^* = 0$ para cualquier solución óptima, prevalecen rendimientos constantes a escala.

$\mathbf{k}_0^* < 0$ para todas las soluciones óptimas, prevalecen rendimientos decrecientes a escala.

La ecuación (3.10) puede expresarse en forma matricial como:

$$\max_{\boldsymbol{\mu}, \boldsymbol{\delta}, \mathbf{k}} w_0 = \boldsymbol{\mu}^\top \mathbf{y}_0 + \mathbf{k}_0 \quad \text{sujeto a} \quad \begin{cases} \boldsymbol{\delta}^\top \mathbf{x}_0 = 1 \\ \boldsymbol{\mu}^\top \mathbf{Y} + \mathbf{k}_0 \leq \boldsymbol{\delta}^\top \mathbf{X} \\ \boldsymbol{\mu}^\top, \boldsymbol{\delta}^\top \geq \mathbf{I}_\varepsilon \\ \mathbf{k}_0 \text{ no restringida} \end{cases} \quad (3.11)$$

El objetivo del modelo orientado a entradas de la ecuación (3.11) es encontrar el hiperplano que permaneciendo sobre o por encima de todas las DMUs minimiza la distancia horizontal desde el hiperplano a la DMU en evaluación.

La DMU evaluada será calificada eficiente si $w_0^* = 1$ y existe al menos un óptimo ($\boldsymbol{\mu}^* > 0, \boldsymbol{\delta}^* > 0$), mientras que el valor de \mathbf{k}^* se usa para identificar el tipo de rendimiento a escala en el que opera la DMU evaluada.

Si se desean construir modelos que supongan rendimientos no crecientes a escala o rendimientos no decrecientes a escala, solo debe modificarse la restricción asociada a \mathbf{k} , tal que: $\mathbf{k}_0 \leq 0$ y $\mathbf{k}_0 \geq 0$, respectivamente.

3.5. Modelo DEA aditivo

Los modelos aditivos [Charnes et al.1985] combinan orientaciones a entradas y salidas en un único modelo. El modelo aditivo considerando rendimientos variables a escala puede

formularse como:

$$\max_{\lambda, s^+, s^-} (\mathbf{Is}^+ + \mathbf{Is}^-) \quad \text{sujeto a} \quad \begin{cases} \lambda \mathbf{Y} - \mathbf{s}^+ = \mathbf{y}_0 \\ -\lambda \mathbf{X} - \mathbf{s}^- = -\mathbf{x}_0 \\ 1\lambda = 1 \\ \lambda, s^+, s^- \geq 0 \end{cases} \quad (3.12)$$

En la ecuación (3.12) todas las ineficiencias son capturadas en las variables \mathbf{s}^+ y \mathbf{s}^- .

3.6. Discusión y consideraciones acerca de DEA

- DEA caracteriza cada una de las DMUs mediante una única puntuación de eficiencia relativa.
- Al proyectar cada DMU ineficiente sobre la envolvente eficiente se destacan áreas de mejora para cada una de las DMU.
- DEA no requiere especificar modelos estadísticos ni hacer inferencias basadas en el análisis de residuos y coeficientes de los parámetros.
- DEA permite contemplar el uso de variables exógenas
- Es posible considerar variables categóricas en el análisis.
- A través de la técnica DEA es posible emplear múltiples entradas y múltiples salidas expresadas en distintas unidades de medida.
- Al ser una técnica no paramétrica no supone ninguna forma funcional de relación entre las entradas y las salidas, ni supone una distribución de eficiencia.
- Al evaluar la eficiencia relativa de una DMU, DEA considera las condiciones más favorables, en tal sentido, si la DMU es ineficiente no puede considerarse que sea a causa de los pesos asignados. Sin embargo, es posible que una DMU sea catalogada como eficiente más por el ajuste de los pesos que por una eficiencia intrínseca.
- Una de las mayores críticas de DEA es que se trata de una aproximación determinista y no tiene en cuenta las influencias sobre el proceso productivo de carácter aleatorio, ni la incertidumbre.
- La precisión de los resultados de eficiencia depende de la exactitud de las medidas de las entradas y las salidas consideradas, así, cuando no hay certeza en los datos empleados la DMU puede llegar a ser catalogada como eficiente o como ineficiente si una verdadera justificación para ello.
- DEA es sensible a la existencia de observaciones extremas y toda desviación respecto de la frontera es tratada como ineficiencia, lo que puede derivar en sobre-estimación de la misma.

- Una gran cantidad de DMUs pueden ser caracterizadas como eficientes a menos que la suma de la cantidad de entradas y salidas sea pequeña en comparación con el número de observaciones (DMUs), en tal caso se aconseja tener al menos el doble de observaciones con respecto a la e suma de cantidad de entradas y salidas.
- La omisión de una entrada o de una salida importante puede generar resultados sesgados.
- DEA no ofrece una medida de eficiencia absoluta, no permite comparar la DMU con un valor máximo teórico de eficiencia.
- La mayor parte de las aplicaciones con DEA se solucionan empleando la forma envolvente, debido a que el modelo multiplicativo tiene un número de restricciones igual a $n + 1$, mientras que el envolvente tiene $s + m$ restricciones, y en la práctica es usual que el número total de DMUs sea mucho mayor que la suma del número de entradas y salidas, es decir, $n \gg s + m$.
- Al elegir el modelo DEA, en aplicaciones en que las entradas no están completamente bajo el control del gestor, los modelos orientados a salidas son adecuados. Por otra parte, si los resultados del proceso (salidas) son decididos por los objetivos de los gestores y no a partir de mejores prácticas observadas, es preferible elegir modelos orientados a entradas.
- Cuando el rendimiento de las DMUs a comparar no depende de la escala de operación, escoger como tipología del modelo rendimientos constantes a escala puede ser adecuado.
- A diferencia de los modelos DEA-CCR, en los que se obtiene una medida de eficiencia técnica global, los modelos DEA-BCC obtienen medidas de eficiencia técnica pura, es decir, se trata de medidas de eficiencia técnica netas de cualquier efecto de escala.
- **Un problema con la fijación de referencias para la mejora en DEA (ver la sección 3.3.1) es que sólo es posible encontrar proyecciones sobre la frontera eficiente, lo cual para las DMUs (empresas) ineficientes puede llegar a ser inalcanzable. En estos casos sería preferible tener la posibilidad de hallar puntos intermedios que mejoren el valor de eficiencia, sin que necesariamente se tenga que alcanzar el máximo de eficiencia técnica.**

Capítulo 4

Análisis de correlación canónica

Cuando un proceso puede ser descrito por dos conjuntos de variables correspondientes a dos aspectos diferentes, analizar las relaciones entre estos dos puntos de vista del mismo proceso puede mejorar la comprensión del sistema subyacente [Uurtio et al.2018]. En este contexto, una relación es una proyección de las observaciones correspondientes a una variables de un punto de vista a las observaciones correspondientes a una variable del otro punto de vista. Estas relaciones pueden ser determinadas por medio de métodos de correlación canónica.

El análisis de correlación canónica (CCA - *Canonical Correlation Analysis*) es un método estadístico multivariado de dos vistas. En el análisis estadístico multivariado los datos corresponden con múltiples variables medidas sobre un conjunto de individuos. Para CCA, las variables de una observación pueden ser divididas en dos conjuntos que pueden ser vistos como las dos vistas de los datos.

Sean las vistas a y b , para los datos, denotadas por las matrices $\mathbf{X}_a \in \mathbb{R}^{n \times p}$ y $\mathbf{X}_b \in \mathbb{R}^{n \times q}$. Los vectores fila $\mathbf{x}_a^k \in \mathbb{R}^p$ y $\mathbf{x}_b^k \in \mathbb{R}^q$ para $k = 1, 2, \dots, n$ denotan los conjuntos de observaciones multivariadas empíricas en \mathbf{X}_a y \mathbf{X}_b , respectivamente. Se asume que las observaciones son muestreadas conjuntamente de una distribución multivariada normal. Los vectores columna $\mathbf{a}_i \in \mathbb{R}^n$ para $i = 1, 2, \dots, p$ y $\mathbf{b}_j \in \mathbb{R}^n$ para $j = 1, 2, \dots, q$ denotan los vectores de variables de las n observaciones. El producto interno entre 2 vectores es denotado por $\langle \mathbf{a}, \mathbf{b} \rangle$ o por $\mathbf{a}^\top \mathbf{b}$. Sin pérdida de generalidad, se puede asumir que las variables tienen media cero y varianza uno. El principal objetivo en CCA es extraer las relaciones lineales entre las variables \mathbf{X}_a y \mathbf{X}_b .

Se pueden considerar las siguientes transformaciones

$$\mathbf{X}_a \mathbf{w}_a = \mathbf{z}_a \quad \text{y} \quad \mathbf{X}_b \mathbf{w}_b = \mathbf{z}_b,$$

donde $\mathbf{w}_a \in \mathbb{R}^p$, $\mathbf{z}_a \in \mathbb{R}^n$, $\mathbf{w}_b \in \mathbb{R}^q$ y $\mathbf{z}_b \in \mathbb{R}^n$. Las matrices \mathbf{X}_a y \mathbf{X}_b representan transformaciones lineales de las posiciones \mathbf{w}_a y \mathbf{w}_b sobre las imágenes \mathbf{z}_a y \mathbf{z}_b en el espacio \mathbb{R}^n . Las posiciones \mathbf{w}_a y \mathbf{w}_b son conocidas como vectores canónicos de pesos y las imágenes \mathbf{z}_a y \mathbf{z}_b son llamadas variaciones canónicas o puntuaciones. Las restricciones de CCA en las proyecciones son que la posición de los vectores de las imágenes \mathbf{z}_a y \mathbf{z}_b son de norma unitaria y que se minimiza el ángulo de cierre, $\theta \in \left[0, \frac{\pi}{2}\right]$, entre \mathbf{z}_a y \mathbf{z}_b . El coseno del ángulo, también llamado correlación canónica, entre las imágenes está dado

por $\cos(\mathbf{z}_a, \mathbf{z}_b) = \langle \mathbf{z}_a, \mathbf{z}_b \rangle / (\|\mathbf{z}_a\| \|\mathbf{z}_b\|)$ y debido a la restricción de norma unitaria se tiene que $\cos(\mathbf{z}_a, \mathbf{z}_b) = \langle \mathbf{z}_a, \mathbf{z}_b \rangle$. De hecho, el principio básico de CCA es hallar dos posiciones $\mathbf{w}_a \in \mathbb{R}^p$ y $\mathbf{w}_b \in \mathbb{R}^q$ que después de las transformaciones lineales $\mathbf{X}_a \in \mathbb{R}^{n \times p}$ y $\mathbf{X}_b \in \mathbb{R}^{n \times q}$ sean mapeadas sobre una bola unitaria n -dimensional y ubicadas de tal forma que se maximice el coseno del ángulo entre los vectores posición de las imágenes $\mathbf{z}_a \in \mathbb{R}^n$ y $\mathbf{z}_b \in \mathbb{R}^n$.

Las imágenes \mathbf{z}_a y \mathbf{z}_b de las posiciones \mathbf{w}_a y \mathbf{w}_b que resultan en el ángulo más pequeño, θ_1 , determinan la primera correlación canónica que es igual a $\cos \theta_1$. El ángulo más pequeño está dado por

$$\cos \theta_1 = \max_{\mathbf{z}_a, \mathbf{z}_b} \langle \mathbf{z}_a, \mathbf{z}_b \rangle \quad \text{sujeto a} \quad \begin{cases} \|\mathbf{z}_a\|_2 = 1 \\ \|\mathbf{z}_b\|_2 = 1 \end{cases} \quad (4.1)$$

El máximo obtenido es para las imágenes \mathbf{z}_a^1 y \mathbf{z}_b^1 . El par de imágenes \mathbf{z}_a^2 y \mathbf{z}_b^2 , que tiene el segundo ángulo de cierre más pequeño θ_2 , son halladas en los complementos ortogonales de \mathbf{z}_a^1 y \mathbf{z}_b^1 . El proceso se repite hasta que no hayan más pares por ser encontrados. Así, los r ángulos $\theta_r \in [0, \frac{\pi}{2}]$ para $r = 1, 2, \dots, q$ cuando $p > q$, pueden ser hallados de forma recursiva como:

$$\cos \theta_r = \max_{\mathbf{z}_a, \mathbf{z}_b} \langle \mathbf{z}_a^r, \mathbf{z}_b^r \rangle \quad \text{sujeto a} \quad \begin{cases} \|\mathbf{z}_a\|_2 = 1 & \|\mathbf{z}_b\|_2 = 1 \\ \langle \mathbf{z}_a^r, \mathbf{z}_a^j \rangle = 0 & \langle \mathbf{z}_b^r, \mathbf{z}_b^j \rangle = 0 \quad \forall j \neq r \end{cases}$$

donde $j, r = 1, 2, \dots, \min(p, q)$.

El número de correlaciones canónicas, r , corresponde con la dimensión de CCA, la cual puede ser vista como el número de patrones que pueden ser extraídos de los datos.

Las posiciones \mathbf{w}_a y \mathbf{w}_b y las imágenes \mathbf{z}_a y \mathbf{z}_b pueden encontrarse resolviendo un problema estándar de valores propios [Urtio et al.2018], obteniéndose:

$$\begin{aligned} \frac{1}{\rho} \mathbf{C}_{ba} \mathbf{C}_{aa}^{-1} \mathbf{C}_{ab} \mathbf{w}_b - \rho \mathbf{C}_{bb} \mathbf{w}_b &= 0 \\ \mathbf{w}_a &= \frac{\mathbf{C}_{aa}^{-1} \mathbf{C}_{ab} \mathbf{w}_b}{\rho} \end{aligned} \quad (4.2)$$

donde $\mathbf{C}_{ab} = \frac{1}{n-1} \mathbf{X}_a^\top \mathbf{X}_b$ es la matriz de covarianza muestral, $\mathbf{C}_{aa} = \frac{1}{n-1} \mathbf{X}_a^\top \mathbf{X}_a$ y $\mathbf{C}_{bb} = \frac{1}{n-1} \mathbf{X}_b^\top \mathbf{X}_b$ son las matrices de varianza empírica y ρ corresponde con el valor propio de la descomposición. Además,

$$\begin{aligned} \mathbf{z}_a &= \mathbf{X}_a \mathbf{w}_a \\ \mathbf{z}_b &= \mathbf{X}_b \mathbf{w}_b \end{aligned} \quad (4.3)$$

Así, de las ecuaciones (4.2) y (4.3) se calculan las posiciones y las imágenes, respectivamente. Se debe tener claridad que cada par \mathbf{w}_a y \mathbf{w}_b tiene su respectivo par de imágenes \mathbf{z}_a y \mathbf{z}_b asociadas, y que la cantidad de pares está dada por r (dimensión CCA).

El par de vectores posición que tienen imágenes sobre la bola unitaria con un ángulo de cierre mínimo corresponden al modelo de correlación canónica obtenido a partir de los datos de entrenamiento. Las entradas de estos vectores de posición transmiten las relaciones entre las variables (obtenidas a partir de la distribución muestral).

Las entradas de los vectores posición \mathbf{w}_a y \mathbf{w}_b pueden ser usadas con el propósito de analizar las relaciones lineales entre las variables. La relación lineal correspondiente al valor de la correlación canónica es hallado entre las entradas que son de mayor valor.

Adicionalmente, un modelo estadístico puede ser validado en términos de su significancia estadística, probando las hipótesis de la siguiente ecuación (4.4):

$$H_0 : \min(p, q) = k \quad \text{vs} \quad H_1 : \min(p, q) > k, \quad (4.4)$$

donde $k = 0, 1, \dots, p$ cuando $p < q$. Si la hipótesis H_0 es rechazada para $j = 0, 1, \dots, k - 1$ y, en vez, aceptada para $H_1 : \min(p, q) > k - 1$ el número de correlaciones canónicas estadísticamente significativas puede ser estimado como k .

Para la prueba de hipótesis el estadístico L_k de Bartlett-Lawley, de la ecuación (4.5), puede ser aplicado:

$$L_k = - \left(n - k - \frac{1}{2}(p + q + 1) + \sum_{j=1}^k r_j^{-2} \right) \ln \left(\prod_{j=k+1}^{\min(p, q)} (1 - r_j^2) \right), \quad (4.5)$$

donde r_j denota la j -ésima correlación canónica. La distribución nula asintótica de L_k es la distribución χ^2 con $(p - k)(q - k)$ grados de libertad. De hecho, primero se prueba que no haya correlación canónica entre las dos vistas, si se rechaza tal hipótesis H_0 , se continúa probando que exista una correlación canónica. En caso que todos los patrones canónicos sean estadísticamente significativos la hipótesis $H_0 : \min(p, q) = k - 1$ será rechazada, y se concluirá que el número de correlaciones canónicas estadísticamente significativas es k .

Capítulo 5

Evaluación comparativa por proyección de distancia mínima

Como se ha mencionado anteriormente, es un problema vigente la identificación e interpretación acerca de qué variables de entrada y en qué medida deben ser intervenidas, en condiciones prácticas, para obtener mejores resultados sobre las variables de salida, mejorando de esta forma el nivel de eficiencia de la empresa.

También se ha señalado que un problema con la fijación de referencias para la mejora empleando DEA (ver la sección 3.3.1) es que la identificación de las variables relevantes (y sus valores de cambio) sólo puede hacerse bajo la hipótesis que la empresa logra la eficiencia técnica relativa máxima, es decir, sólo es posible encontrar proyecciones sobre la frontera eficiente, lo cual, en condiciones prácticas, cotidianas y reales, para las DMUs (empresas) ineficientes, puede llegar a ser inalcanzable. En estos casos sería preferible tener la posibilidad de hallar puntos intermedios que mejoren el valor de eficiencia, sin que necesariamente se tenga que alcanzar el máximo de eficiencia técnica.

En tal sentido, en este capítulo se plantea una metodología para establecer los valores de cambio para cada una de las variables (relevantes) de una empresa ineficiente, de manera que la empresa ineficiente pueda asemejarse a un conglomerado de empresas eficientes.

En este caso, sea $\mathbf{X} \in \mathbb{R}^{n \times p}$ una matriz con las variables observadas de entrada para el conjunto de las DMUs, donde el vector $\mathbf{x}_i \in \mathbb{R}^p$ para $i = 1, 2, \dots, n$ denota las entradas multivariadas de la i -ésima DMU. Sea $\mathbf{Y} \in \mathbb{R}^{n \times q}$ una matriz con las variables observadas de salida para el conjunto de las mismas DMUs donde $\mathbf{y}_i \in \mathbb{R}^q$ son las salidas multivariadas de la i -ésima DMU.

Una vez aplicado el análisis envolvente de datos (DEA) se obtendrá el vector de etiquetas, $\mathbf{u} \in \mathbb{R}^{n \times 1}$, que contiene la clasificación de cada una de las n DMUs en las categorías ineficiente y eficiente, asignándoles una representación numérica de -1 y 1 , respectivamente. Adicionalmente, el resultado del análisis de correlación canónica (CCA) dará información acerca de cuales variables de entrada son relevantes. La matriz de entradas \mathbf{X} será transformada en la matriz $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times m}$, donde $m \leq p$, de tal manera que $\tilde{\mathbf{X}}$ contiene las variables de entrada relevantes observadas para todas las DMUs.

Consideremos la submatriz $\tilde{\mathbf{X}}_E \in \mathbb{R}^{h \times m}$ que contiene h filas de $\tilde{\mathbf{X}}$ que corresponden con las empresas eficientes, es decir, asociadas a $u_j = 1 \forall j$. A partir de las observaciones de la matriz $\tilde{\mathbf{X}}_E$ se podrá construir una envolvente convexa. Además, está el vector $\mathbf{t} \in \mathbb{R}^m$

que representa a cualquiera de las empresas ineficientes, asociada con $u_t = -1$.

Realizando una combinación lineal de las DMUs de $\tilde{\mathbf{X}}_E$ se puede obtener un punto $\tilde{\mathbf{t}} \in \mathbb{R}^m$ que vive sobre la envolvente convexa, y que representará la proyección de la DMU ineficiente sobre la envolvente convexa construida a partir de las DMUs eficientes, el cual será la versión eficiente de la empresa ineficiente, y que está dado por la ecuación (5.1):

$$\tilde{\mathbf{t}} = \tilde{\mathbf{X}}_E^\top \boldsymbol{\alpha} \quad (5.1)$$

donde $\boldsymbol{\alpha} \in \mathbb{R}^h$ es un vector de ponderación para la combinación lineal de las DMUs.

Ahora, es también necesario garantizar que la proyección eficiente $\tilde{\mathbf{t}}$, de la DMU ineficiente \mathbf{t} , genere la menor cantidad de cambios entre los valores de \mathbf{t} y $\tilde{\mathbf{t}}$. Por tal motivo nosotros planteamos un método que hemos denominado **proyección de distancia mínima** (MDM - *Minimum Distance Mapping*), el cual busca la mínima distancia entre el punto \mathbf{t} y la envolvente convexa sobre la cual vive $\tilde{\mathbf{t}}$. Esta condición implica que la DMU ineficiente tendrá que hacer la menor cantidad de cambios posibles en sus variables de entrada para asemejarse al conjunto de empresas eficientes, tal como se puede observar en la Figura 5.1.

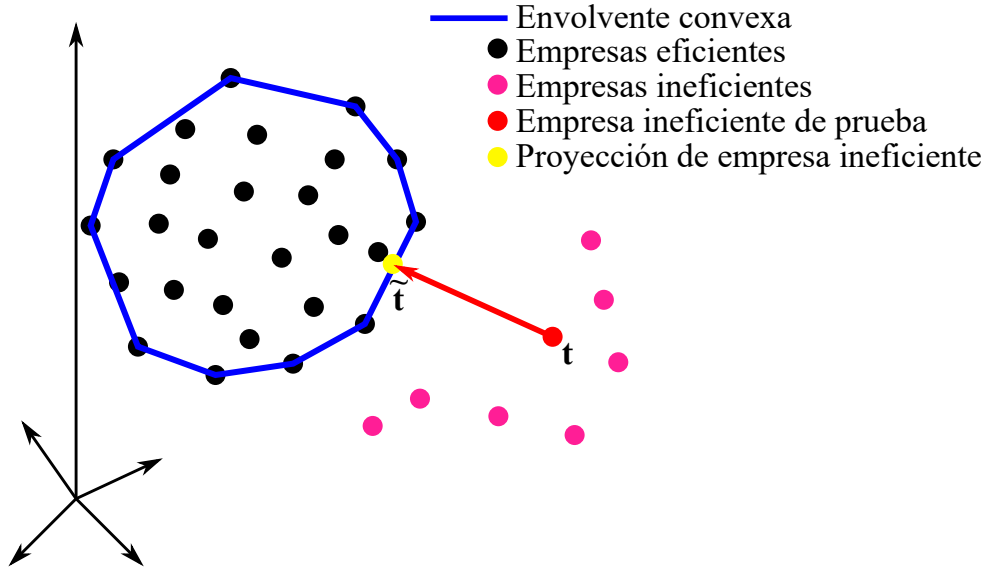


Figura 5.1: Proyección de distancia mínima - MDM

La proyección de distancia mínima (MDM) puede formularse como el siguiente problema de optimización, presentado en la ecuación (5.2):

$$\min_{\boldsymbol{\alpha}} \|\mathbf{t} - \tilde{\mathbf{t}}\|_2 = \min_{\boldsymbol{\alpha}} \|\mathbf{t} - \tilde{\mathbf{X}}_E^\top \boldsymbol{\alpha}\|_2 \quad \text{sujeto a} \quad \begin{cases} \alpha_i \geq 0 \quad \forall i \\ \|\boldsymbol{\alpha}\|_2 = 1 \end{cases} \quad (5.2)$$

La ecuación (5.2) puede reescribirse de forma cuadrática de la siguiente manera:

$$\begin{aligned}
 \min_{\alpha} \left\| \mathbf{t} - \tilde{\mathbf{X}}_E^\top \alpha \right\|_2 &= \min_{\alpha} \left(\mathbf{t} - \tilde{\mathbf{X}}_E^\top \alpha \right)^\top \left(\mathbf{t} - \tilde{\mathbf{X}}_E^\top \alpha \right) \\
 &= \min_{\alpha} \left[\mathbf{t}^\top - \left(\tilde{\mathbf{X}}_E^\top \alpha \right)^\top \right] \left(\mathbf{t} - \tilde{\mathbf{X}}_E^\top \alpha \right) \\
 &= \min_{\alpha} \left(\mathbf{t}^\top - \alpha^\top \tilde{\mathbf{X}}_E \right) \left(\mathbf{t} - \tilde{\mathbf{X}}_E^\top \alpha \right) \\
 &= \min_{\alpha} \left(\mathbf{t}^\top \mathbf{t} - \alpha^\top \tilde{\mathbf{X}}_E \mathbf{t} - \mathbf{t}^\top \tilde{\mathbf{X}}_E^\top \alpha + \alpha^\top \tilde{\mathbf{X}}_E \tilde{\mathbf{X}}_E^\top \alpha \right),
 \end{aligned}$$

y dado que α y \mathbf{t} son vectores columna, se tiene

$$\alpha^\top \tilde{\mathbf{X}}_E \mathbf{t} = \mathbf{t}^\top \tilde{\mathbf{X}}_E^\top \alpha$$

por tanto, la ecuación (5.2) puede reescribirse como en la ecuación (5.3):

$$\min_{\alpha} \left(\mathbf{t}^\top \mathbf{t} - 2\alpha^\top \tilde{\mathbf{X}}_E \mathbf{t} + \alpha^\top \tilde{\mathbf{X}}_E \tilde{\mathbf{X}}_E^\top \alpha \right) \quad \text{sujeto a} \quad \begin{cases} \alpha_i \geq 0 \quad \forall i \\ \|\alpha\|_2 = 1 \end{cases} \quad (5.3)$$

Ahora, haciendo los siguientes reemplazos de variables:

$$\begin{aligned}
 \mathbf{Q} &= \tilde{\mathbf{X}}_E \tilde{\mathbf{X}}_E^\top \\
 \mathbf{c} &= \tilde{\mathbf{X}}_E \mathbf{t} \quad \Rightarrow \quad \mathbf{c}^\top = \mathbf{t}^\top \tilde{\mathbf{X}}_E^\top
 \end{aligned}$$

se tiene

$$\begin{aligned}
 \alpha^\top \tilde{\mathbf{X}}_E \mathbf{t} &= \alpha^\top \mathbf{c} \\
 \left(\alpha^\top \tilde{\mathbf{X}}_E \mathbf{t} \right)^\top &= \left(\alpha^\top \mathbf{c} \right)^\top \\
 \mathbf{t}^\top \tilde{\mathbf{X}}_E^\top \alpha &= \mathbf{c}^\top \alpha
 \end{aligned}$$

con lo cual se obtiene,

$$\min_{\alpha} \left(\alpha^\top \mathbf{Q} \alpha - 2\mathbf{c}^\top \alpha + \mathbf{t}^\top \mathbf{t} \right) \quad \text{sujeto a} \quad \begin{cases} \alpha_i \geq 0 \quad \forall i \\ \|\alpha\|_2 = 1 \end{cases} \quad (5.4)$$

y como $\mathbf{t}^\top \mathbf{t}$ no depende de α , al derivar e igualar a cero para resolver el problema de minimización, este término desaparece, quedando:

$$\min_{\alpha} \left(\alpha^\top \mathbf{Q} \alpha - 2\mathbf{c}^\top \alpha \right) \quad \text{sujeto a} \quad \begin{cases} \alpha_i \geq 0 \quad \forall i \\ \|\alpha\|_2 = 1 \end{cases} \quad (5.5)$$

Finalmente, reescribiendo la ecuación (5.5), se llega a la expresión cuadrática de la ecuación (5.6):

$$\min_{\alpha} \left(\frac{1}{2} \alpha^\top \mathbf{Q} \alpha - \mathbf{c}^\top \alpha \right) \quad \text{sujeto a} \quad \begin{cases} \alpha_i \geq 0 \quad \forall i \\ \mathbf{1}^\top \alpha = 1 \end{cases} \quad (5.6)$$

donde $\mathbf{1}$ es un vector de unos de tamaño $h \times 1$.

La solución del problema cuadrático de la ecuación (5.6) puede ser hallada empleando el método de multiplicadores de Lagrange [Madsen et al.2004], donde la función de Lagrange asociada es:

$$L(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{1}{2} \boldsymbol{\alpha}^\top \mathbf{Q} \boldsymbol{\alpha} - \mathbf{c}^\top \boldsymbol{\alpha} - \sum \lambda (\mathbf{1}^\top \boldsymbol{\alpha} - 1)$$

Tales métodos de optimización cuadrática general se encuentran incorporados en la mayoría de paquetes de software matemáticos.

Capítulo 6

Metodología

Esta investigación presenta una metodología de análisis de relevancia de variables financieras por medio de la combinación de técnicas de análisis de frontera eficiente con técnicas de aprendizaje de máquina, para establecer, cuáles variables financieras y en qué proporción se deben intervenir, en pro de optimizar la eficiencia de las empresas.

En general, en este capítulo se explica cómo la propuesta metodológica permite conseguir el objetivo planteado. Tal como se muestra en la Figura 6.1, la propuesta metodológica se estructura en cuatro etapas: preproceso, evaluación de eficiencia, relevancia de variables y evaluación comparativa.

6.1. Preproceso

Con el fin de poder realizar las operaciones necesarias a lo largo de la metodología, en primera instancia se ha establecido una etapa de preproceso de los datos, para la cual se procede de la siguiente manera: 1) Remoción de observaciones con valores faltantes, 2) Remoción de variables con más de 20 % de valores iguales a cero, 3) Centralización de los datos, 4) Estandarización de los datos, y 5) Ajuste a no-negatividad en las variables.

El componente de remoción de observaciones con valores faltantes consiste en eliminar, de los datos, las empresas para las cuales no haya valor registrado de alguna de las 11 variables de entrada o de alguna de las 3 variables de salida.

Posteriormente, para cada conjunto de empresas en evaluación, se cuantifica en cada variable la cantidad de valores iguales a cero y se calcula la proporción de valores cero con relación a la cantidad total de empresas en el grupo, en caso que para alguna variable esta proporción supere el 20 %, se elimina la variables correspondiente.

Sea $\hat{\mathbf{X}}$ la matriz con los datos de entrada y $\hat{\mathbf{Y}}$ la matriz con los datos de salida, donde las filas corresponden a las observaciones (empresas o DMUs) y las columnas a las variables. Considere $\hat{\mathbf{X}}$ y $\hat{\mathbf{Y}}$ depurados para las observaciones y las variables como se mencionó en los párrafos anteriores, el paso siguiente es realizar la centralización y estandarización de los datos, es decir, llevarlos a media cero y varianza uno, tal que:

$$\frac{\hat{\mathbf{X}} - \mu_{\hat{\mathbf{X}}}}{\sigma_{\hat{\mathbf{X}}}} \quad \frac{\hat{\mathbf{Y}} - \mu_{\hat{\mathbf{Y}}}}{\sigma_{\hat{\mathbf{Y}}}} \quad (6.1)$$

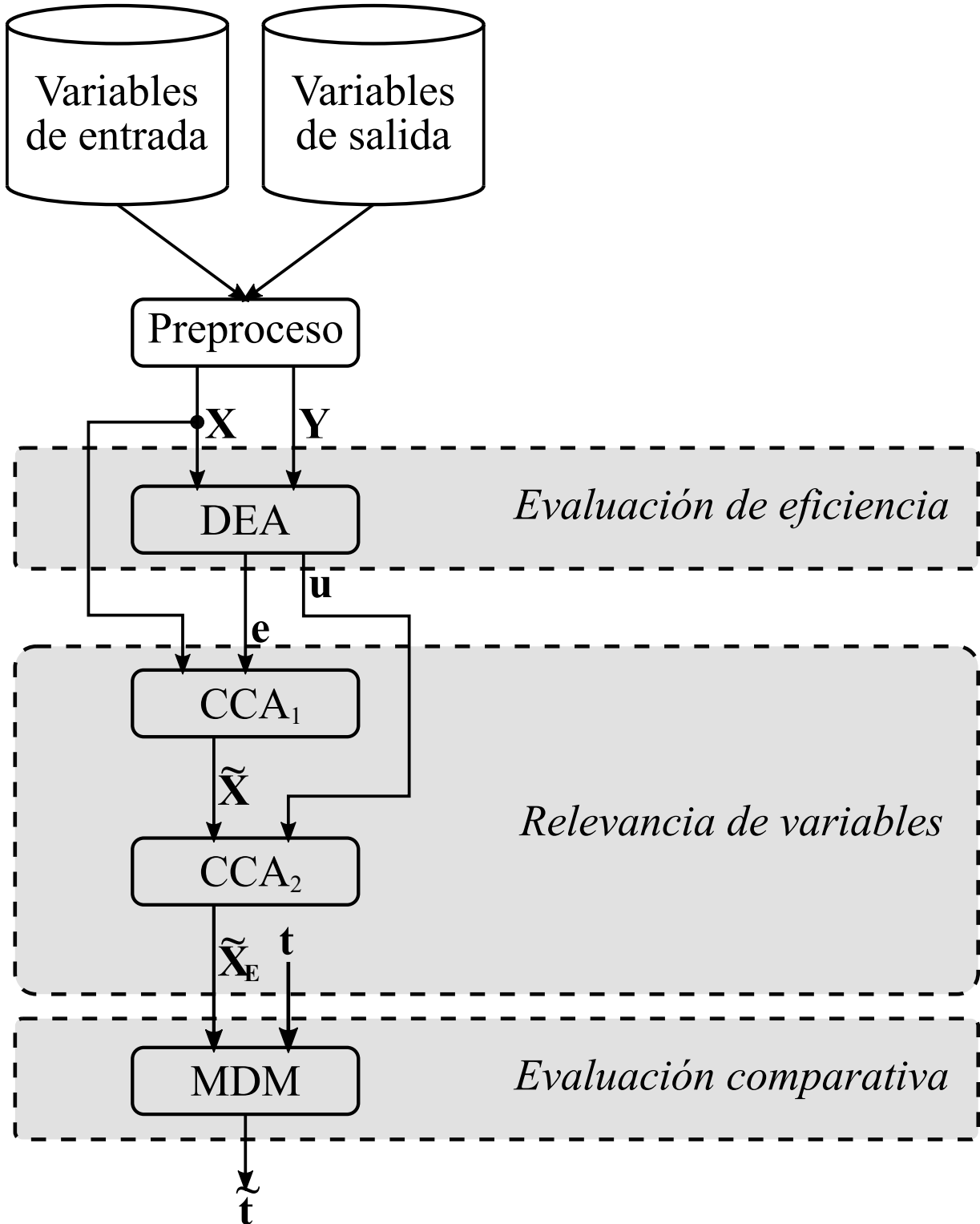


Figura 6.1: Esquema metodológico

donde

$$\mu_{\hat{\mathbf{X}}} = \frac{1}{n} \sum_{i=1}^n \hat{x}_i$$

es el vector de medias, y

$$\sigma_{\hat{\mathbf{x}}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\hat{\mathbf{x}}_i - \boldsymbol{\mu}_{\hat{\mathbf{x}}})^2}$$

es el vector que contiene las desviaciones estándar de cada variable, siendo $\hat{\mathbf{x}}_i$ el vector fila asociado a la i -ésima empresa. Semejantes son las expresiones asociadas a $\hat{\mathbf{Y}}$.

Finalmente, y dado que el análisis envolvente de datos (DEA) requiere que los valores de las observaciones sean todos no negativos, se aplica la aproximación clásica invariante a traslación que se presenta en [Pastor and Ruiz2007] para transformar los datos. Como resultado de esta etapa se obtiene $\mathbf{X} \in \mathbb{R}^{n \times p}$ correspondiente a la matriz de entradas y $\mathbf{Y} \in \mathbb{R}^{n \times q}$ correspondiente a la matriz de salidas.

6.2. Evaluación de eficiencia

Se tiene $\mathbf{X} \in \mathbb{R}^{n \times p}$ que es una matriz con las variables observadas de entrada para el conjunto de las DMUs, donde el vector $\mathbf{x}_i \in \mathbb{R}^p$ para $i = 1, 2, \dots, n$ denota las entradas multivariadas de la i -ésima DMU, y $\mathbf{Y} \in \mathbb{R}^{n \times q}$ que es una matriz con las variables observadas de salida para el conjunto de las mismas DMUs donde $\mathbf{y}_i \in \mathbb{R}^q$ son las salidas multivariadas de la i -ésima DMU.

Con los datos de entrada \mathbf{X} y salida \mathbf{Y} para el conjunto de DMUs, se calcula la eficiencia relativa de cada una de las DMUs empleando análisis envolvente de datos (DEA). Con propósitos exclusivamente comparativos se implementan cinco modelos diferentes del análisis envolvente de datos: DEA-CCR orientado a entradas, DEA-CCR orientado a salidas, DEA-BCC orientado a entradas, DEA-BCC orientado a salidas, y DEA-Aditivo.

Como se ha mencionado, DEA permite calcular la eficiencia relativa de las DMUs, obteniéndose el vector $\mathbf{e} \in \mathbb{R}^n$, en el cual el elemento e_i para $i = 1, 2, \dots, n$ corresponde con la eficiencia relativa de la i -ésima DMU. Adicionalmente, conociendo el vector de eficiencias, es posible establecer las empresas que son eficientes y las empresas que no lo son, dando lugar al vector de etiquetas de eficiencia \mathbf{u} con n elementos, de modo que $u_i \in \{-1, 1\}$ para $i = 1, 2, \dots, n$, es decir:

$$u_i = \begin{cases} 1 & \text{si } e_i \geq 1 \\ -1 & \text{si } e_i < 1 \end{cases}$$

6.3. Relevancia de variables

En esta etapa se tienen dos prioridades: 1) determinar cuáles son las variables de entrada a partir de $\mathbf{X} \in \mathbb{R}^{n \times p}$ que tienen mayor nivel de relación con la eficiencia $\mathbf{e} \in \mathbb{R}^n$ previamente calculada, y 2) con las variables de entrada seleccionadas $\tilde{\mathbf{X}}$ y sus correspondientes etiquetas de eficiencia \mathbf{u} , se pretende establecer el conjunto de empresas modelo (eficientes y cuasi-eficientes) que serán el referente de comparación para las empresas ineficientes.

En primera instancia, a través del análisis de correlación canónica (CCA) se pueden calcular los vectores canónicos de pesos \mathbf{w}_a y \mathbf{w}_b (ver Capítulo 4), los cuales explican, para este caso, las relaciones lineales entre las variables de entrada y la eficiencia, donde

los mayores pesos implican mayor relación. Sin embargo, para dar validez a las relaciones halladas, es necesario que el valor de la correlación canónica sea lo más alto posible, y que se pueda establecer significancia estadística. En tal sentido, se ordenan de forma descendente los pesos al interior de \mathbf{w}_a , dando orden de prioridad a las variables de entrada, y se procede a calcular nuevamente la correlación canónica para el caso en que en la entrada solo se tiene la variable de mayor importancia y posteriormente agregando las demás variables hasta encontrar que la correlación no incrementa, es decir, que se adicionarán las variables de entrada necesarias (en orden de importancia) para que el nivel de correlación se haga máximo o al menos no haya un incremento considerable al agregar más variables. El conjunto de variables resultante corresponde con la matriz $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times m}$, donde $m \leq p$, de tal manera que $\tilde{\mathbf{X}}$ contiene las variables de entrada relevantes observadas para todas las DMUs. Puede darse la situación en que sea necesario agregar todas las variables de entrada para lograr la mayor correlación, y se tendrá que $\mathbf{X} = \tilde{\mathbf{X}}$. Una vez establecida $\tilde{\mathbf{X}}$ se verifica la significancia estadística de la correlación canónica entre esta matriz y la eficiencia.

Para el segundo propósito, se realiza CCA entre la matriz de entradas priorizadas $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times m}$ y el vector de etiquetas de eficiencia \mathbf{u} de n elementos, para obtener una representación alternativa de los datos dada por el vector de imágenes $\mathbf{z}_a \in \mathbb{R}^n$. A partir de las proyecciones \mathbf{z}_a se ajusta una función de distribución de probabilidad normal para el grupo de empresas eficientes y una función de distribución de probabilidad normal para el grupo de empresas ineficientes, adicionalmente, se determina un umbral, que permite seleccionar cuántas y cuáles de las DMUs eficientes conforman el grupo modelo o grupo referencia de comparación. Lo anterior se presenta gráficamente en la Figura 6.2.

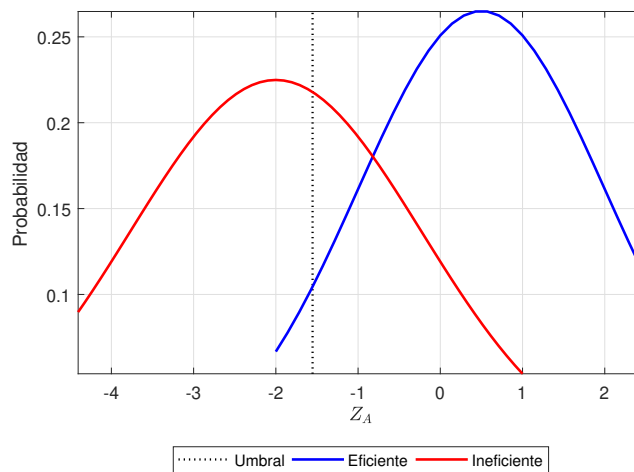


Figura 6.2: Distribución de probabilidad de las imágenes de CCA

El resultado de esta etapa es la matriz de datos $\tilde{\mathbf{X}}_E \in \mathbb{R}^{h \times m}$ que contiene h filas de $\tilde{\mathbf{X}}$ que corresponden con las empresas que conforman el grupo modelo o referencia de comparación.

6.4. Evaluación comparativa por proyección de distancia mínima

Esta etapa se desarrolla con el fin de comparar alguna (cualquiera) de las empresas ineficientes contra el grupo de empresas modelo, de manera que sea posible establecer, para cada una de las variables de entrada priorizadas, los cambios necesarios que deben realizarse, para que así, la empresa ineficiente se asemeje al conjunto de empresas modelo.

En este sentido, empleando el método propuesto denominado proyección de distancia mínima (MDM), se compara la DMU ineficiente $\mathbf{t} \in \mathbb{R}^m$ contra el conjunto de empresas modelo dado por $\tilde{\mathbf{X}}_E \in \mathbb{R}^{h \times m}$, tal como se explica en el Capítulo 5.

Como resultado de esta etapa se obtiene $\tilde{\mathbf{t}} \in \mathbb{R}^m$, que es la proyección de \mathbf{t} sobre la envolvente convexa dada por el grupo de empresas modelo. Esto significa que $\tilde{\mathbf{t}}$ es la versión eficiente de la empresa ineficiente, y además, asegura la menor cantidad de cambios posibles en las variables de entrada priorizadas de la empresa ineficiente para asemejarse al conjunto de empresas modelo.

Capítulo 7

Experimentación

El universo de aplicación de la propuesta está enmarcado a empresas pequeñas Colombianas con ánimo de lucro, con datos de estados financieros de un corte particular. Con el fin de que haya consistencia en la información y sea posible realizar comparaciones justas entre las empresas que se analicen, es necesario que las empresas pertenezcan a un mismo sector económico y que sean de igual tamaño empresarial (dado por el total de activos).

Los ejemplos que se presentan en este capítulo están contruidos a partir de datos financieros disponibles para el año 2015 en el sistema de información y reporte empresarial (SIREM) de la Superintendencia de Sociedades. El SIREM registra información del estado de resultados y del estado de situación financiera (balance general). En particular, las variables financieras consideradas para el análisis fueron establecidas con base en su disponibilidad para la gran mayoría de las empresas presentes en la base de datos; específicamente fueron seleccionadas 11 variables de entrada y 3 variables de salida como se detalla en la Tabla 7.1.

Variables de entrada	Variables de salida
Activos corrientes	Utilidad bruta
Activos no corrientes	Utilidad operacional
Total de activos	Ganancias y pérdidas
Pasivos corrientes	
Pasivos no corrientes	
Total pasivos	
Ingresos operacionales	
Costo de ventas	
Gastos operacionales de administración	
Gastos operacionales de ventas	
Intereses	

Tabla 7.1: Variables de entrada y salida

7.1. Datos empleados

En este trabajo se presentan los resultados de la metodología propuesta a partir de dos conjuntos de datos diferentes. El primer conjunto de datos es de empresas de tamaño pequeño pertenecientes al sector económico: Fabricación de productos de caucho y plástico (códigos CIIU Sección C División 22), el segundo conjunto de datos es de empresas de tamaño pequeño pertenecientes al sector económico: Confección de prendas de vestir (códigos CIIU Sección C División 14). Sobre estos datos se procede a eliminar las observaciones que tienen valores faltantes en las variables de entrada y salida.

En la Tabla 7.2 se presenta la cantidad de empresas pequeñas de cada uno de los dos sectores económicos que están en la base de datos del SIREM 2015, y la cantidad de empresas restantes luego de eliminar las observaciones con datos faltantes.

Código	Sector	No. empresas	No. empresas depuradas
C22	Fabricación de productos de caucho y plástico	110	$n_{C22} = 99$
C14	Confección de prendas de vestir	134	$n_{C14} = 119$

Tabla 7.2: Cantidad de empresas pequeñas por sector

Adicionalmente, son eliminadas las variables para las cuales el 20 % o más de sus valores son iguales a cero. En la Tabla 7.3 se presenta la remoción de variables para cada uno de los sectores económicos.

Código	Sector	Variables removidas
C22	Fabricación de productos de caucho y plástico	Pasivos no corrientes
C14	Confección de prendas de vestir	Pasivos no corrientes

Tabla 7.3: Variables removidas en conjunto de datos

En síntesis, hay $p = 10$ variables de entrada y $q = 3$ variables de salida para ambos sectores económicos que se están evaluando (ver Tabla 7.4). Adicionalmente, los estimadores de media, desviación estándar, máximo y mínimo para las variables de entrada y salida de los sectores C22 y C14 están dados en las Tablas 7.5 y 7.6, respectivamente.

Código	No. variables entrada	No. variables salida	No. observaciones
C22	$p = 10$	$q = 3$	$n = 99$
C14	$p = 10$	$q = 3$	$n = 119$

Tabla 7.4: Dimensiones de los conjuntos de datos

Variable	Media	Desv.	Mín.	Máx.
(1) Activos corrientes	1.245.850	679.539	74.318	2.955.189
(2) Activos no corrientes	658.009	587.021	0	2.348.405
(3) Total activos	1.903.858	793.064	326.599	3.193.160
(4) Pasivos corrientes	826.905	658.136	24.373	4.594.589
(5) Total pasivos	1.078.683	932.002	93.218	7.689.336
(6) Ingresos operacionales	2.682.749	2.147.335	0	16.320.826
(7) Costo de ventas	1.940.889	1.706.493	0	13.564.745
(8) Gastos operacionales de administración	401.661	341.534	0	2.427.997
(9) Gastos operacionales de ventas	203.543	263.805	0	1.586.727
(10) Intereses	39.851	45.678	0	208.002
(S1) Utilidad bruta	741.860	653.592	-887.113	3.440.732
(S2) Utilidad operacional	136.656	298.576	-1.000.039	1.369.515
(S3) Ganancias y pérdidas	29.242	316.814	-1.811.252	937.355

Tabla 7.5: Descripción de las variables en el sector C22. Valores expresados en miles.

Variable	Media	Desv.	Mín.	Máx.
(1) Activos corrientes	1.422.055	730.749	4.564	3.204.144
(2) Activos no corrientes	352.438	374.164	0	1.894.865
(3) Total activos	1.774.493	772.583	542.870	3.208.193
(4) Pasivos corrientes	758.748	533.475	9.640	2.874.032
(5) Total pasivos	905.210	561.735	39.906	3.062.971
(6) Ingresos operacionales	3.027.745	4.786.796	0	38.572.395
(7) Costo de ventas	1.959.732	3.037.366	0	30.639.965
(8) Gastos operacionales de administración	406.454	484.020	0	4.268.381
(9) Gastos operacionales de ventas	490.806	2.211.989	0	23.613.730
(10) Intereses	62.235	273.913	0	2.961.282
(S1) Utilidad bruta	1.068.013	2.615.113	-267.808	27.193.947
(S2) Utilidad operacional	170.753	304.554	-688.164	2.476.889
(S3) Ganancias y pérdidas	59.269	346.126	-2.978.475	1.701.253

Tabla 7.6: Descripción de las variables en el sector C14. Valores expresados en miles.

7.2. Resultados

En esta sección se presentan los resultados de las etapas metodológicas: evaluación de eficiencia, relevancia de variables, y evaluación comparativa, calculadas para los conjuntos de empresas explicados en la Sección 7.1.

7.2.1. Resultados de eficiencia

Como se menciona anteriormente, para calcular la eficiencia de las empresas se implementaron, con propósitos comparativos, los modelos DEA-CCR orientado a entradas, DEA-CCR orientado a salidas, DEA-BCC orientado a entradas, DEA-BCC orientado a salidas, y DEA-Aditivo.

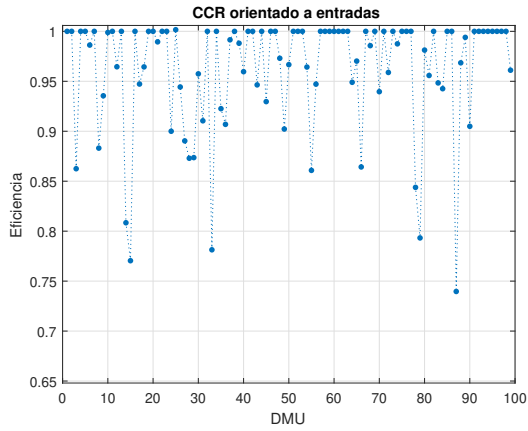
En la Figura 7.1 se presentan los resultados de eficiencia para el conjunto de datos C22. En general, se observa como los métodos CCR (Figuras 7.1(a) y 7.1(b)), reporta más cantidad de empresas ineficientes que el método BCC (Figuras 7.1(c) y 7.1(d)). En tal sentido, es coherente que la eficiencia calculada a través del método modelo BCC sea generalmente mayor que la calculada por el modelo CCR, debido a la consideración de los retornos a escala del modelo BCC, y adicionalmente porque el modelo CCR asume una eficiencia global mientras que el BCC asume solo la eficiencia técnica. Por otra parte, las DMUs eficientes del modelo BCC orientado a entradas son también eficientes en el modelo Aditivo y viceversa, sin embargo para las DMUs ineficientes, el nivel de eficiencia calculado difiere entre ambos modelos, reflejándose con mayor severidad la ineficiencia en el modelo Aditivo.

En la Figura 7.2 se presentan los resultados de eficiencia para el conjunto de datos C14. En este caso, primero se verifica el hecho de que el modelo BCC calcula mayores eficiencias que el modelo CCR, y adicionalmente, entre los modelos comparados, el modelo Aditivo (Figura 7.2(e)) es el que presenta DMUs con menores valores de eficiencia.

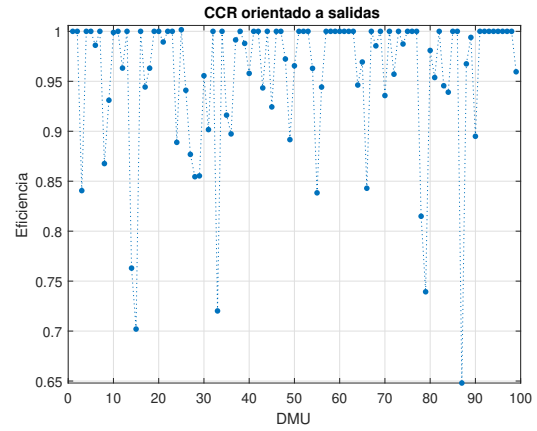
En las Tablas 7.7 y 7.8 (para los conjuntos de datos C22 y C14 respectivamente) se presentan matrices que señalan la cantidad y el porcentaje de DMUs eficientes por cada uno de los modelos y la cantidad de DMUs eficientes comunes entre los mismos. Es de destacar que alrededor de la mitad de las DMUs son señaladas como eficientes en cada modelo, lo cual lleva a pensar que no hay grandes sesgos por atipicidad en la eficiencia y que el conjunto de DMUs eficientes (en cada sector) tiene un comportamiento similar. Coincidentalmente, tanto para el conjunto C22 como para el conjunto C14, al cruzar todos los modelos, se encuentran 36 DMUs eficientes en común.

	CCR-IO	CCR-OO	BCC-IO	BCC-OO	Aditivo
CCR-IO	46 (46.5 %)	40 (40.4 %)	45 (45.5 %)	41 (41.4 %)	46 (46.5 %)
CCR-OO		42 (42.4 %)	41 (41.4 %)	39 (39.4 %)	42 (42.4 %)
BCC-IO			58 (58.6 %)	48 (48.5 %)	57 (57.6 %)
BCC-OO				49 (49.5 %)	49 (49.5 %)
Aditivo					58 (58.6 %)

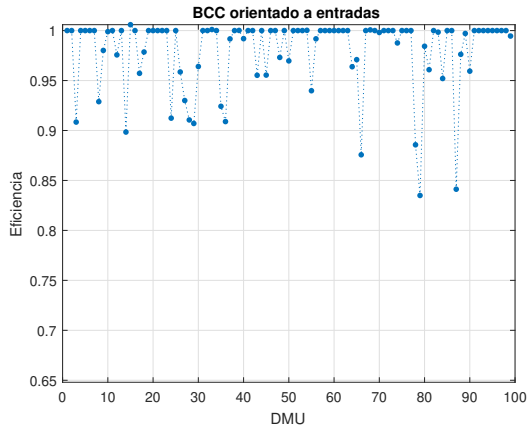
Tabla 7.7: Cantidad y porcentaje de DMUs eficientes, para C22



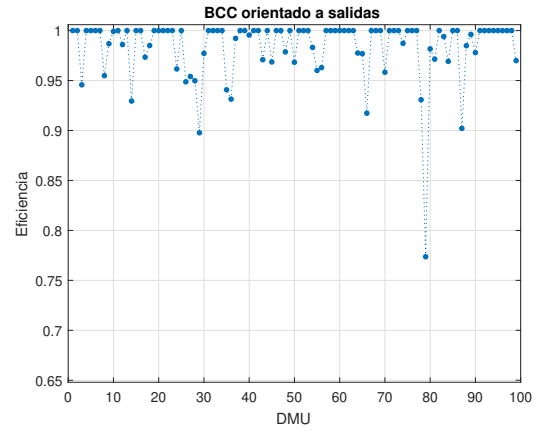
(a) CCR orientado a entradas



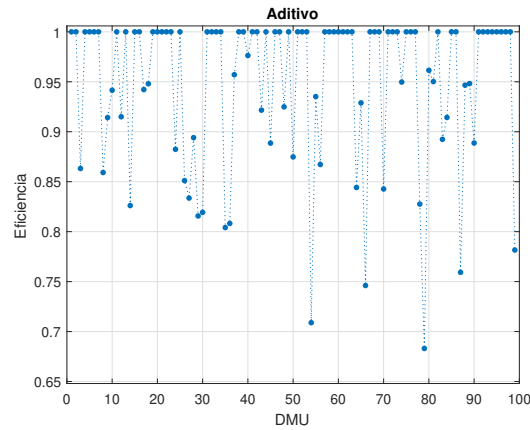
(b) CCR orientado a salidas



(c) BCC orientado a entradas



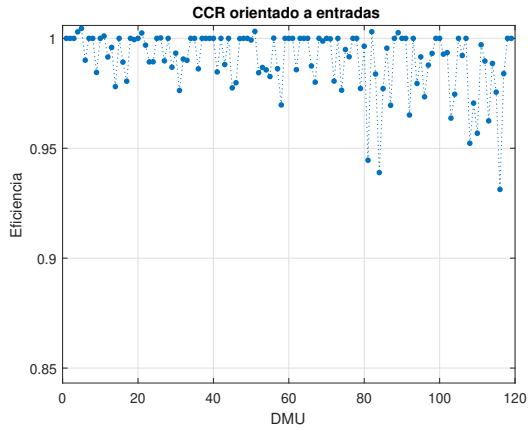
(d) BCC orientado a salidas



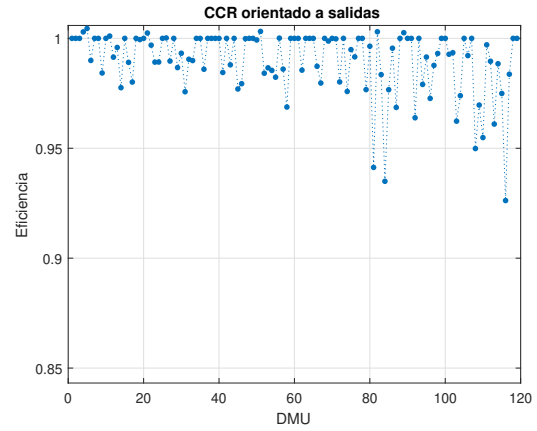
(e) Aditivo

Figura 7.1: Evaluación de eficiencia para el conjunto de datos C22

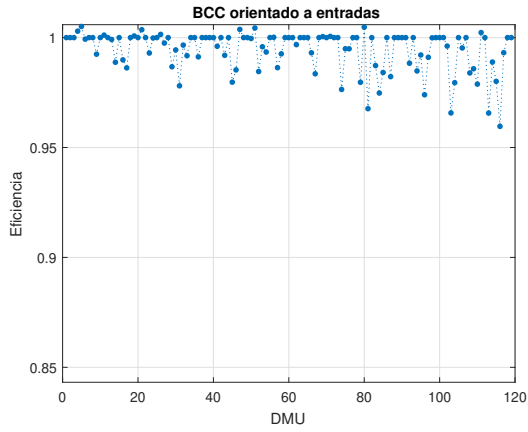
Para el conjunto de datos C22, los modelos que más DMUs eficientes determinaron son: BCC-IO (orientado a entradas) y Aditivo, mientras que el que menos DMUs eficientes halló



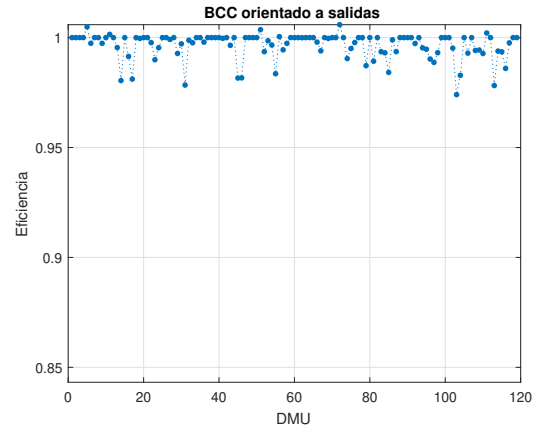
(a) CCR orientado a entradas



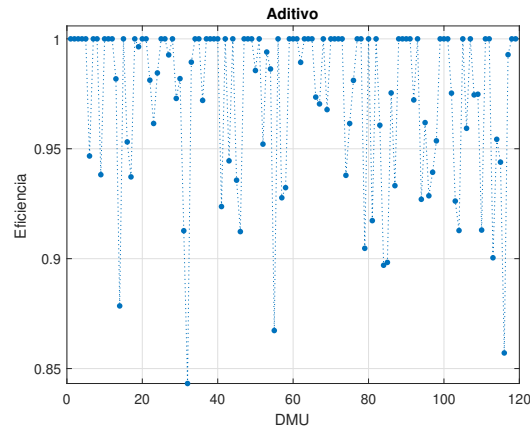
(b) CCR orientado a salidas



(c) BCC orientado a entradas



(d) BCC orientado a salidas



(e) Aditivo

Figura 7.2: Evaluación de eficiencia para el conjunto de datos C14

fue CCR-OO (orientado a salidas). Similarmente, para el conjunto de datos C14, el modelo BCC-IO (orientado a entradas) fue el que más DMUs eficientes identifica, mientras que el

	CCR-IO	CCR-OO	BCC-IO	BCC-OO	Aditivo
CCR-IO	51 (42.9 %)	45 (37.8 %)	49 (41.2 %)	43 (36.1 %)	51 (42.9 %)
CCR-OO		45 (37.8 %)	43 (36.1 %)	38 (31.9 %)	45 (37.8 %)
BCC-IO			62 (52.1 %)	49 (41.2 %)	57 (47.9 %)
BCC-OO				52 (43.7 %)	51 (42.9 %)
Aditivo					59 (49.6 %)

Tabla 7.8: Cantidad y porcentaje de DMUs eficientes, para C14

modelo CCR-OO (orientado a salidas) es el que señala menos DMUs eficientes.

7.2.2. Resultados de relevancia de variables

Inicialmente se determinan cuáles variables de entrada tienen mayor nivel de relación con la eficiencia calculada, a través de la evaluación de los vectores de posición, este cálculo se realiza para cada modelo de eficiencia. Adicionalmente, se verifica el nivel de correlación entre las variables y la eficiencia, agregando las variables en el orden de relevancia dado por el vector posición.

Como ejemplo, consideremos el caso de los resultados de eficiencia del modelo CCR-IO para el conjunto C22. El vector de posición \mathbf{w}_a se grafica en la Figura 7.3(a), donde se puede apreciar que las variables de entrada tienen el siguiente orden de relevancia: (6) Ingresos operacionales, (7) Costo de ventas, (9) Gastos operacionales de ventas, (8) Gastos operacionales de administración, (10) Intereses, (2) Activos no corrientes, (1) Activos corrientes, (4) Pasivos corrientes, (5) Total pasivos, (3) Total activos.

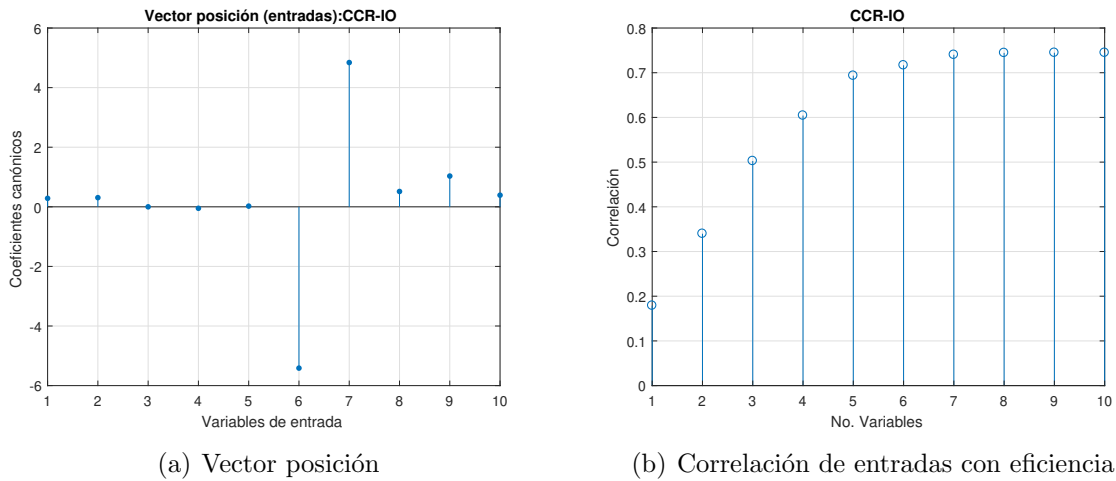


Figura 7.3: Relación de entradas con eficiencia para CCR-IO, conjunto C22

Una vez determinado el orden de relevancia de las variables, se procede a calcular el nivel de correlación del conjunto de variables de entrada con respecto a la eficiencia (Figura 7.3(b)), para ello se van agregando las variables en el orden determinado, de tal manera que en cada iteración la correlación evaluada cuenta con una variable más. En este caso se

inicia con solo la variable (6) Ingresos operacionales, y se finaliza con el conjunto completo de 10 las variables. De la Figura 7.3(b) se puede observar que, en general, a medida que se adicionan variables el nivel de correlación entre estas variables y la eficiencia, aumenta. En este trabajo, el criterio de parada para seleccionar la cantidad de variables de entrada que tienen mayor relación con la eficiencia es encontrar el codo en la curva de correlación (ver Figura 7.4), de tal forma que se obtenga una adecuada compensación entre el número de variables y la correlación obtenida.

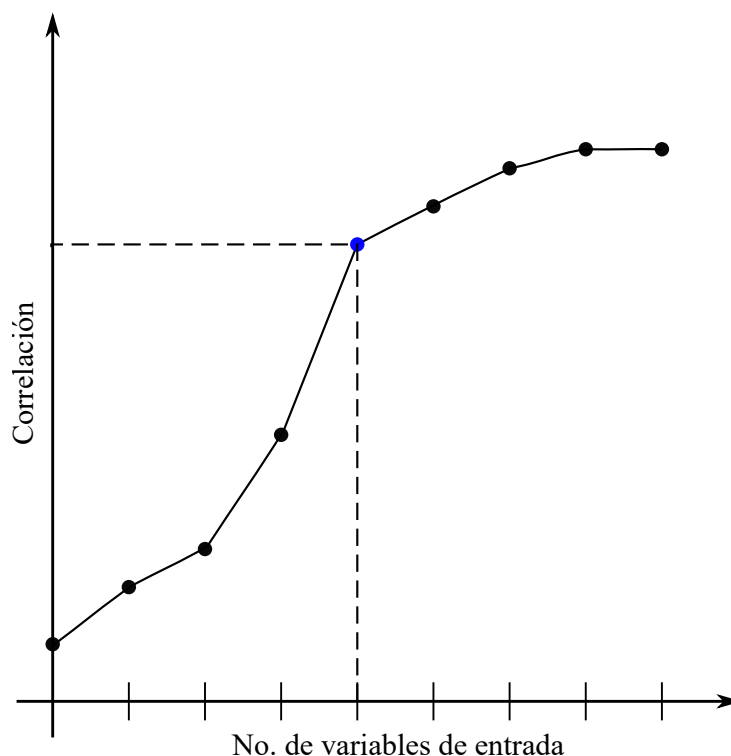
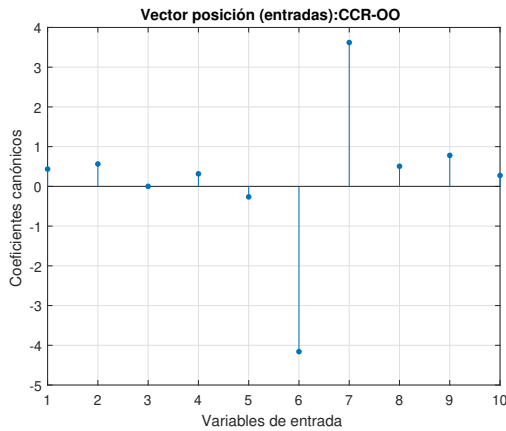


Figura 7.4: Selección del número de variables relevantes

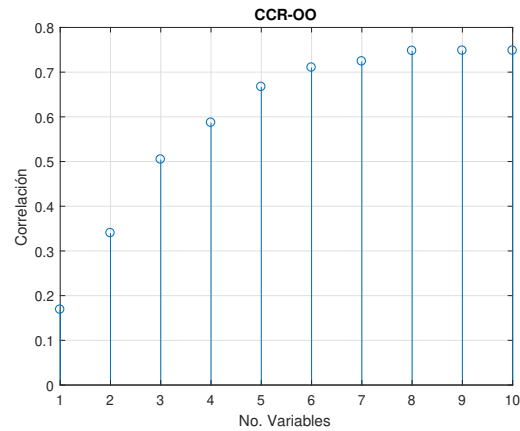
Este mismo análisis se realizó para los demás modelos de cálculo de eficiencia CCR-OO (Figura 7.5), BCC-IO (Figura 7.6), BCC-OO (Figura 7.7) y Aditivo (Figura 7.8).

Para el conjunto de datos C22, en la Tabla 7.9 se presenta, para cada uno de los cinco modelos de eficiencia, cuáles y cuántas fueron las variables relevantes seleccionadas, el valor de correlación alcanzado entre estas variables y la eficiencia, y el valor- p que verifica la significancia estadística con el 95 % de confianza.

Iguales cálculos se realizaron para los datos C14. En la Tabla 7.10 se presenta cuántas y cuáles fueron las variables relevantes, la correlación y el valor- p . Para todos los casos calculados, las variables predominantes en el análisis de relevancia son: Ingresos operacionales y Costos de ventas. Lo cual es un resultado coherente, dado que las variables de salida con las cuales se calculó la eficiencia se construyen principalmente con estas variables relevantes. También es de notar que para ambos ejemplos, en los cinco modelos, siempre fue descartada la variable Total de activos, lo cual puede explicarse por el hecho de que esta variable está principalmente explicada a través de la variable Activos corrientes, la cual si es señalada como relevante en gran parte de los casos. Al comparar las variables

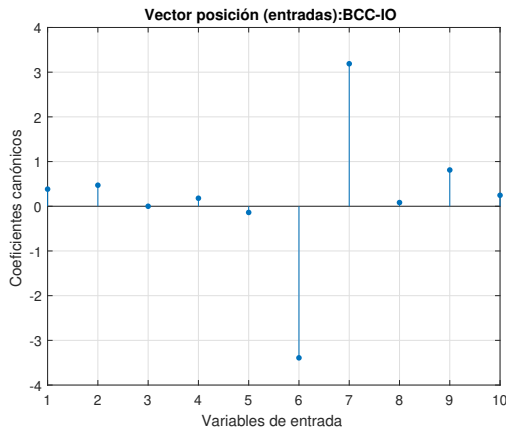


(a) Vector posición

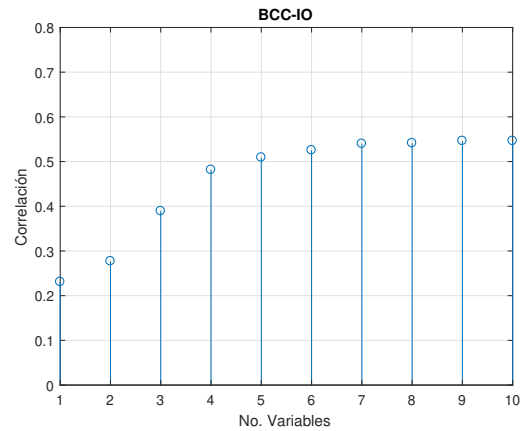


(b) Correlación de entradas con eficiencia

Figura 7.5: Relación de entradas con eficiencia para CCR-OO, conjunto C22



(a) Vector posición



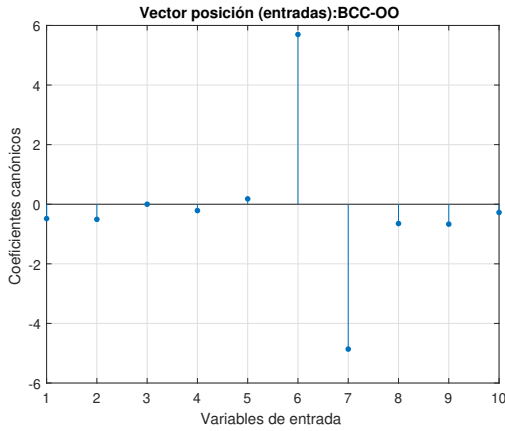
(b) Correlación de entradas con eficiencia

Figura 7.6: Relación de entradas con eficiencia para BCC-IO, conjunto C22

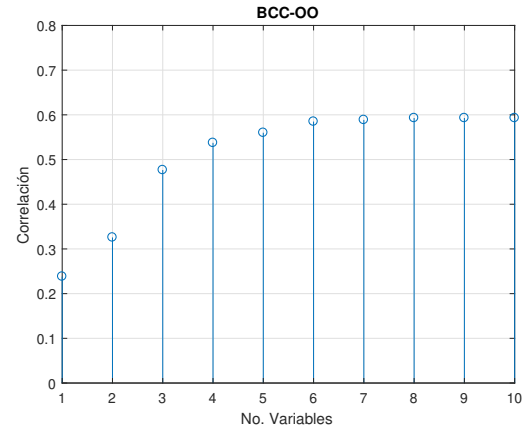
seleccionadas para el sector C22 con las variables relevantes del sector C14 se puede ver que: en el conjunto de datos C22 se selecciona, en general, un mayor número de variables que el sector C14, además para el conjunto de datos C14 es preponderante el Total de pasivos, variable no señalada como relevante de manera común para el conjunto de datos C22, lo cual puede tener explicación en el hecho de que el nivel de endeudamiento de las empresas de C22 es mayor que el de las empresas de C14, y por tanto la correcta administración de la deuda puede ser un elemento diferenciador para ambos sectores.

Tanto para el caso del conjunto de datos C22 como para el conjunto de datos C14, las pruebas de significancia estadística confirman que los patrones de relación entre variables identificados por el modelo de correlación canónica no se presentan por cuestiones del azar.

Ahora, para determinar el conjunto de las DMUs modelo (ejemplares), que sirven como referente de comparación para las DMUs ineficientes, se realiza nuevamente el cálculo de CCA, pero esta vez entre la matriz de variables de entrada relevantes y el vector de

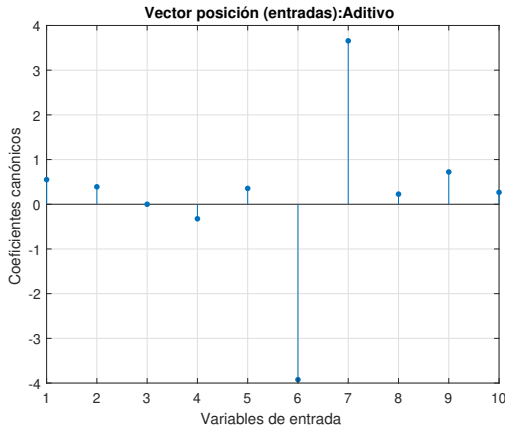


(a) Vector posición

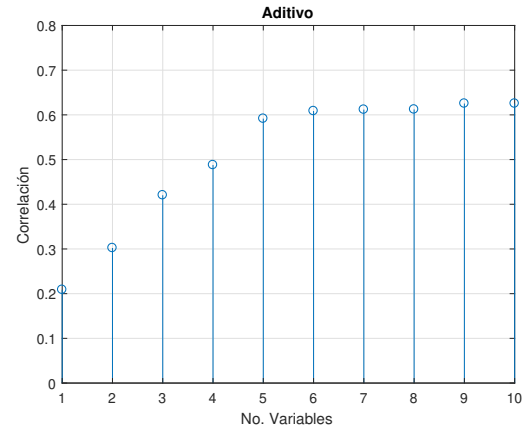


(b) Correlación de entradas con eficiencia

Figura 7.7: Relación de entradas con eficiencia para BCC-OO, conjunto C22



(a) Vector posición



(b) Correlación de entradas con eficiencia

Figura 7.8: Relación de entradas con eficiencia para Aditivo, conjunto C22

etiquetas de eficiencia, con lo cual se obtiene una representación alternativa de los datos dada por el vector de imágenes \mathbf{z}_a con las cuales ajusta una función de distribución de probabilidad normal para el grupo de empresas eficientes y una función de distribución de probabilidad normal para el grupo de empresas ineficientes, adicionalmente, se fija empíricamente un umbral, que permite seleccionar cuántas y cuáles de las DMUs eficientes conforman el grupo referencia de comparación.

Para fijar el parámetro de umbral entre las distribuciones, en este trabajo, se busca reducir el área de traslape entre las distribuciones eficientes e ineficientes, propendiendo porque las muestras que se seleccionen finalmente como las de referencia tengan la menor probabilidad de traslape con el grupo ineficiente, en tal sentido, al rededor de la media del grupo ineficiente se captura se calcula el intervalo que contenga al menos el 40 % del total de las DMUs ineficientes. Los límites de este intervalo sirven como punto extremo sobre la función de distribución de las eficientes, y permite determinar hasta dónde se

Modelo	No. variables	Variables seleccionadas	Correl.	valor- p
CCR-IO	7	Ingresos operacionales, Costo de ventas, Gastos operacionales de ventas, Gastos operacionales de administración, Intereses, Activos no corrientes, Activos corrientes	0,7402	$2,0636 \times 10^{-13}$
CCR-OO	8	Ingresos operacionales, Costo de ventas, Gastos operacionales de ventas, Activos no corrientes, Gastos operacionales de administración, Activos corrientes, Pasivos corrientes, Intereses	0,7477	$2,9127 \times 10^{-13}$
BCC-IO	7	Ingresos operacionales, Costo de ventas, Gastos operacionales de ventas, Activos no corrientes, Activos corrientes, Intereses, Pasivos corrientes	0,5399	$3,7069 \times 10^{-5}$
BCC-OO	6	Ingresos operacionales, Costo de ventas, Gastos operacionales de ventas, Gastos operacionales de administración, Activos no corrientes, Activos corrientes	0,5849	$6,072 \times 10^{-7}$
Aditivo	6	Ingresos operacionales, Costo de ventas, Gastos operacionales de ventas, Activos corrientes, Activos no corrientes, Total pasivos	0,6086	$9,3155 \times 10^{-8}$

Tabla 7.9: Relevancia de variables para el conjunto de datos C22

toman DMUs para el conjunto de referencia de las eficientes, ver Figura 7.9. Sin embargo, es necesario aclarar que esta etapa es totalmente opcional dentro de la metodología y se realiza con el fin principal de que las muestras ineficientes no vayan a quedar al interior de la envolvente convexa que se crea en la última etapa de la metodología.

Para el caso particular del conjunto de datos C22 las funciones de distribución estimadas y los umbrales determinados se presentan en la Figura 7.10.

Una vez, establecidos los umbrales se realiza la selección de las DMUs eficientes que

Modelo	No. variables	Variables seleccionadas	Correl.	valor- p
CCR-IO	4	Ingresos operacionales, Costo de ventas, Intereses, Total pasivos	0,4921	$2,0234 \times 10^{-6}$
CCR-OO	4	Ingresos operacionales, Costo de ventas, Intereses, Total pasivos	0,4946	$1,6908 \times 10^{-6}$
BCC-IO	4	Ingresos operacionales, Costo de ventas, Gastos operacionales de ventas, Total pasivo	0,4245	$1,3542 \times 10^{-4}$
BCC-OO	4	Ingresos operacionales, Costo de ventas, Intereses, Total pasivos	0,3688	0,0021
Aditivo	6	Ingresos operacionales, Costo de ventas, Gastos operacionales de ventas, Intereses, Gastos operacionales de administración, Total pasivos	0,5729	$3,9988 \times 10^{-8}$

Tabla 7.10: Relevancia de variables para el conjunto de datos C14

sirven como referencia de comparación para las DMUs ineficientes. En el ejemplo del conjunto de datos C22, los resultados de esta selección de DMUs se presentan en la Figura 7.11. En este trabajo en particular, dado el tipo de información que se analiza, se establece que para pertenecer al grupo de empresas modelo los valores de las salidas, de tales empresas, deben ser no negativos. Como se puede observar, los puntos azules corresponden con las DMUs eficientes seleccionadas como referentes para las DMUs ineficientes.

El número de empresas eficientes seleccionadas como referentes para cada modelo de eficiencia y para cada conjunto de datos es presentado en la Tabla 7.11. Sorprende la escasa cantidad de DMUs eficientes seleccionadas para el conjunto de datos C22 con el modelo de eficiencia Aditivo, tan solo 2. Este resultado se puede observar gráficamente en las distribuciones de probabilidad estimadas a partir de las imágenes \mathbf{z}_a para DMUs eficientes e ineficientes (Figura 7.10(e)), en donde es claro como las dos funciones de probabilidad se encuentran solapadas casi en su totalidad.

	CCR-IO	CCR-OO	BCC-IO	BCC-OO	Aditivo
C22	34	30	37	38	2
C14	39	35	49	42	47

Tabla 7.11: Cantidad de DMUs eficientes seleccionadas como referencia

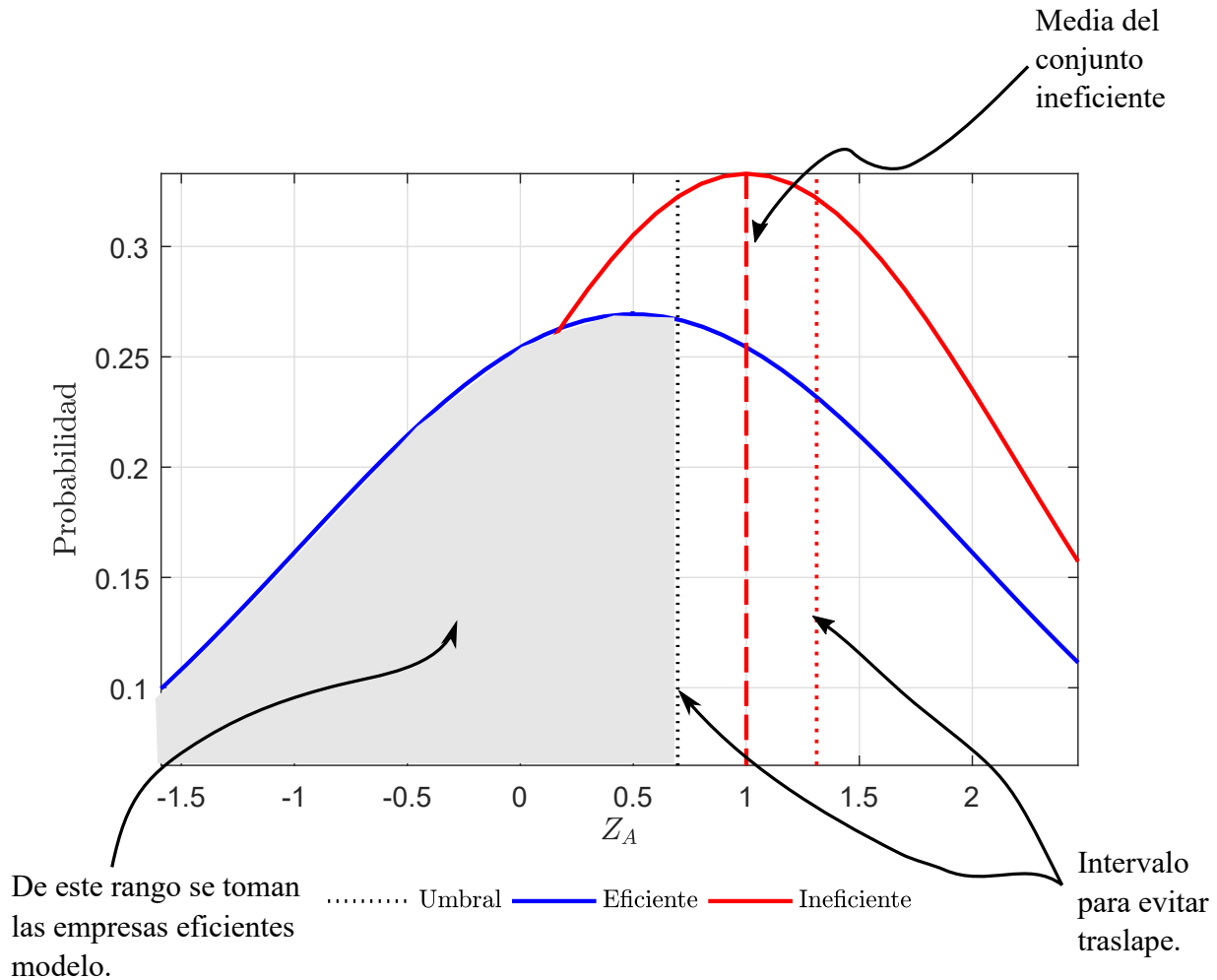
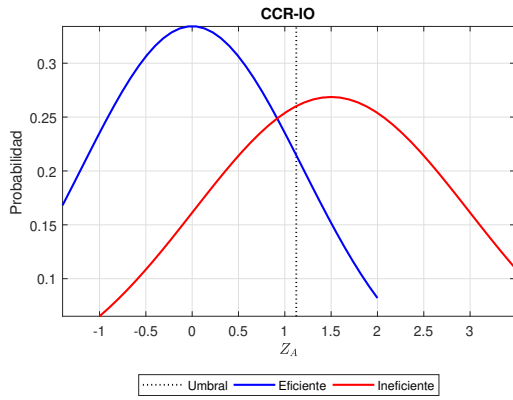


Figura 7.9: Selección de las DMUs eficientes de referencia

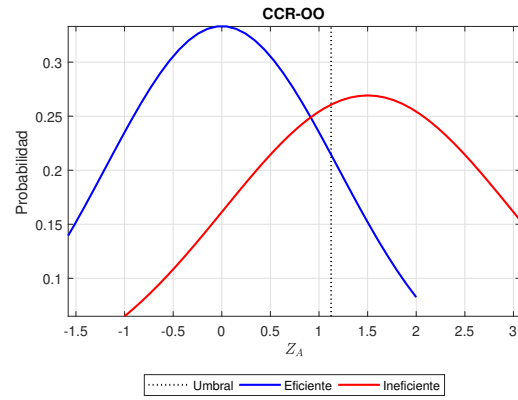
7.2.3. Resultados de evaluación comparativa por proyección de distancia mínima

Una vez seleccionadas las variables relevantes y las DMUs eficientes que sirven como referencia para las ineficientes, se procede a construir la envolvente convexa de las DMUs seleccionadas y posteriormente proyectar las DMUs ineficientes sobre dicha envolvente, donde tal proyección es la versión eficiente de la empresa ineficiente, asegurando la menor cantidad de cambios posibles en las variables relevantes de entrada de la empresa ineficiente para asemejarse al conjunto de empresas modelo.

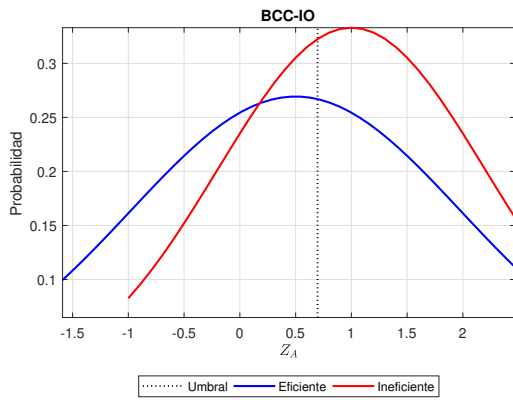
Como ejemplos de aplicación de la evaluación comparativa (*benchmarking*), se han seleccionado dos empresas, una del conjunto de datos C22 y otra del C14. En cada caso las empresas seleccionadas fueron etiquetadas como ineficientes por los cinco modelos de eficiencia analizados. Estas empresas seleccionadas se presentan en la Tabla 7.12. Una vez realizada la proyección de la empresa ineficiente sobre la envolvente convexa se obtienen los valores objetivo en cada una de las variables relevantes, lo cual es presentado para el conjunto de datos C22 en la Tabla 7.13, y para el conjunto de datos C14 en la Tabla 7.14.



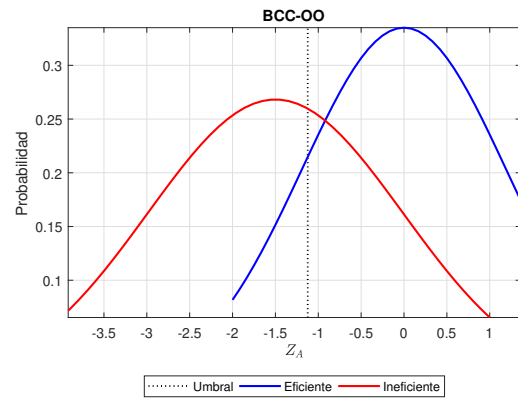
(a) CCR orientado a entradas



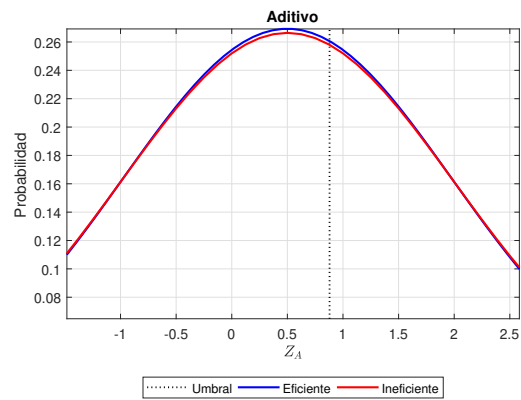
(b) CCR orientado a salidas



(c) BCC orientado a entradas

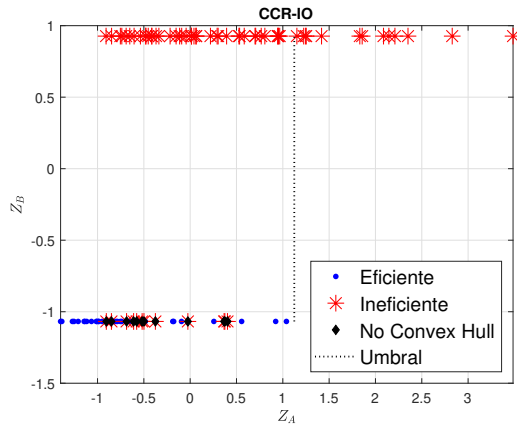


(d) BCC orientado a salidas

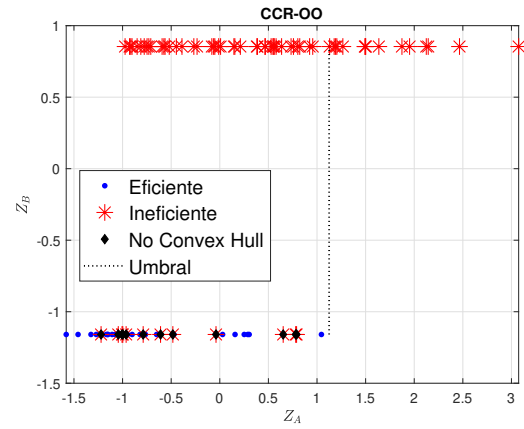


(e) Aditivo

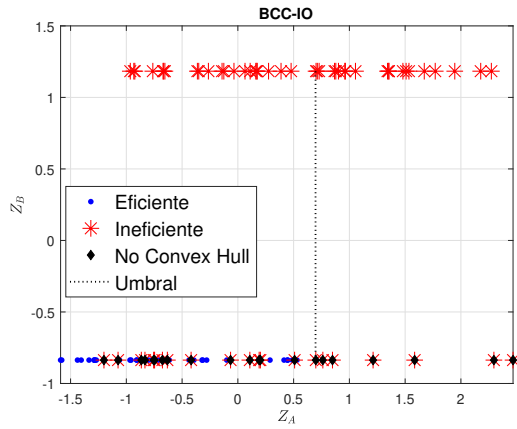
Figura 7.10: Distribuciones estimadas a partir de las imágenes de CCA, conjunto C22



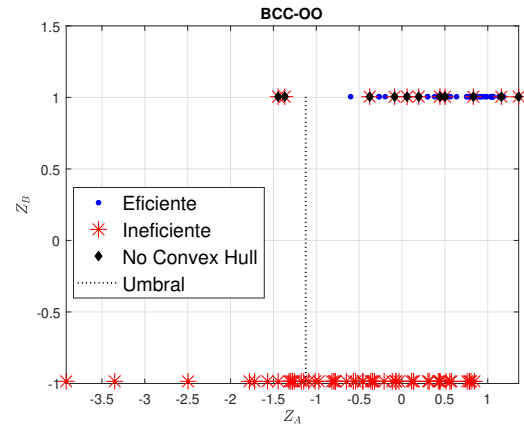
(a) CCR orientado a entradas



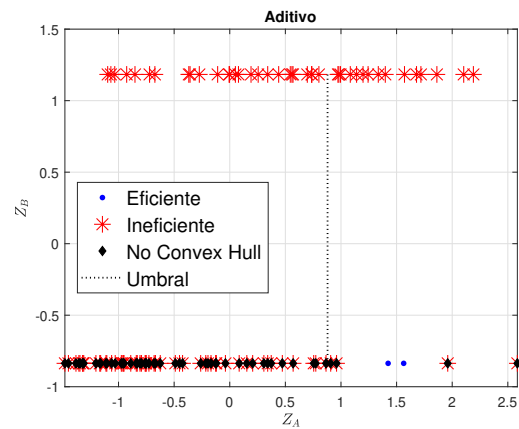
(b) CCR orientado a salidas



(c) BCC orientado a entradas



(d) BCC orientado a salidas



(e) Aditivo

Figura 7.11: DMUs eficientes seleccionadas como referentes, conjunto C22

Variable	C22 Valor	C14 Valor
NIT	800.064.971	800.049.085
*Activo corriente	973.880	1.913.571
*Activo no corriente	1.399.318	532.183
*Total activos	2.373.198	2.445.754
*Pasivo corriente	880.198	1.315.075
*Total pasivos	1.290.768	1.315.075
*Ingresos operacionales	2.710.507	2.625.469
*Costo de ventas	1.901.547	1.813.459
*Gastos operacionales de administración	530.368	622.216
*Gastos operacionales de ventas	159.101	4.240
*Intereses	33.301	61.762
*Utilidad bruta	808.960	812.010
*Utilidad operacional	119.491	185.554
*Ganancias y pérdidas	15.677	86.702

Tabla 7.12: Empresas ineficientes para evaluación comparativa. *Valores en miles

Variables	CCR-IO			CCR-OO			BCC-IO		
	[%]	*Valor	Cambio	[%]	*Valor	Cambio	[%]	*Valor	Cambio
		objetivo			objetivo			objetivo	
Activo corriente	↑ 0,06	974.441,00		↑ 8,25	1.054.207,09		↑ 6,2	1.034.251,64	
Activo no corriente	↓ 0,29	1.395.234,81		↓ 6,75	1.304.813,12		↓ 1,96	1.371.896,38	
Total activos	—			—			—		
Pasivo corriente	—			↓ 9,93	792.810,35		↓ 6,85	819.945,77	
Total pasivos	—			—			—		
Ingresos operacionales	↑ 0,91	2.735.092,78		↑ 3,22	2.797.862,61		↑ 1,33	2.746.509,66	
Costo ventas	↓ 1,42	1.874.589,54		↓ 4,57	1.814.647,07		↓ 1,97	1.864.041,65	
Gastos oper. admin.	↑ 3,88	509.785,95		↑ 8,74	576.730,20		—		
Gastos oper. ventas	↓ 9,06	144.682,23		↓ 37,84	98.897,82		↓ 28,01	114.531,15	
Intereses	↓ 31,75	22.729,25		↓ 95,66	1.446,03		↓ 78,18	7.265,20	
	BCC-OO			Aditivo					
	[%]	*Valor	Cambio	[%]	*Valor	Cambio		*Valor	
		objetivo			objetivo			objetivo	
Activo corriente	↓ 0,24	971.547,41		↓ 10,05	876.047,04				
Activo no corriente	↓ 9,56	1.265.574,55		↓ 54,15	641.543,44				
Total activos	—			—					
Pasivo corriente	—			—					
Total pasivos	—			↓ 36,7	817.010,07				
Ingresos operacionales	↑ 1,09	2.739.947,11		↑ 5,95	2.871.763,99				
Costo ventas	↓ 0,27	1.896.369,69		↑ 9,18	2.076.017,10				
Gastos oper. admin.	↓ 2,1	519.205,87		—					
Gastos oper. ventas	↓ 62,12	60.261,87		↑ 185,24	453.818,61				
Intereses	—			—					

Tabla 7.13: Evaluación comparativa, conjunto C22. *Valores en miles.

Variables	CCR-IO			CCR-OO			BCC-IO		
	[%]	*Valor		[%]	*Valor		[%]	*Valor	
	Cambio	objetivo		Cambio	objetivo		Cambio	objetivo	
Activo corriente	—	—		—	—		—	—	
Activo no corriente	—	—		—	—		—	—	
Total activos	—	—		—	—		—	—	
Pasivo corriente	—	—		—	—		—	—	
Total pasivos	↓ 10,02	1.183.304,05		↓ 0,08	1.314.012,00		↓ 0,01	1.314.987,56	
Ingresos operacionales	↓ 0,16	2.621.318,24		↑ 0,03	2.626.230,85		↑ 0,002	2.625.522,94	
Costo ventas	↑ 1,65	1.843.366,63		↓ 0,02	1.813.117,12		↓ 0,002	1.813.414,83	
Gastos oper. admin.	—	—		—	—		—	—	
Gastos oper. ventas	—	—		—	—		↑ 3,61	4.393,04	
Intereses	↓ 24,85	46.415,65		↓ 22,37	47.947,49		—	—	
	BCC-OO			Aditivo					
	[%]	*Valor		[%]	*Valor				
	Cambio	objetivo		Cambio	objetivo				
Activo corriente	—	—		—	—				
Activo no corriente	—	—		—	—				
Total activos	—	—		—	—				
Pasivo corriente	—	—		—	—				
Total pasivos	↓ 3,45	1.269.716,85		↓ 5,65	1.240.806,23				
Ingresos operacionales	↑ 0,48	2.637.940,93		↑ 1,67	2.669.432,60				
Costo ventas	↓ 0,26	1.808.752,92		↓ 2,32	1.771.439,32				
Gastos oper. admin.	—	—		↓ 12,81	542.499,61				
Gastos oper. ventas	—	—		↓ 40,41	2.526,48				
Intereses	↓ 85,01	9.256,31		↓ 90,82	5.671,40				

Tabla 7.14: Evaluación comparativa, conjunto C14. *Valores en miles.

De los resultados obtenidos para el conjunto de datos C22 (Tabla 7.13) se puede observar que, para todos los modelos, el mayor cambio requerido está asociado a la variable Intereses, la evaluación en los modelos CCR y BCC para esta variable implica una reducción de los intereses, sin embargo el modelo Aditivo solicita un aumento de la misma variable. Esta incongruencia, en el modelo Aditivo, puede ser consecuencia de la escasa cantidad de DMUs eficientes seleccionadas como referencia, solo de dos (Tabla 7.11), y por tanto los resultados no son generalizantes de las condiciones del grupo. Adicionalmente, los modelos CCR y BCC requieren el incremento de la variable Ingresos operacionales y el decremento de los Costo de venta, conservando total coherencia con el propósito del experimento, teniendo en cuenta que estas variables fueron las más relevantes (Tabla 7.9).

Para los resultados del conjunto C14 (Tabla 7.14) se tiene que la variable Intereses fue señalada por 4 modelos (CCR-IO, CCR-OO, BCC-OO y Aditivo) como la que requiere más cambio, en particular debe ser decrementada, además en estos casos la variable Total pasivos también debe decrementarse. Aunque el modelo BCC-IO pareciera presentar requerimientos incoherentes al exigir disminución de los Ingresos operacionales y el aumento en el Costo de venta, tal situación debe revisarse detalladamente pues en realidad los porcentajes de cambio son casi cero, lo que indica que en general esta empresa ya se comporta como el grupo de referencia en las variables indicadas.

7.3. Discusión de los resultados y consideraciones prácticas

- Es fundamental establecer adecuadamente las variables de entrada y salida para capturar una relación subyacente de eficiencia. En tal sentido, es en primera instancia aconsejable establecer la lógica o intencionalidad del análisis. Como criterio amplio y general: DEA minimiza las entradas y maximiza las salidas, en otras palabras, menores niveles de las entradas y mayores niveles de las salidas representan mejores eficiencias [Cook et al.2014]. En este trabajo en particular, la selección de entradas y salidas satisface tal condición lógica.
- En la literatura es usual encontrar que el número de DMUs debe ser al menos el doble del número de entradas más salidas (regla de oro) [Golany and Roll1989], porque a mayor cantidad de observaciones, mayor es la probabilidad de capturar las unidades de alto desempeño que determinan la frontera eficiente. Sin embargo, tal regla no es imperativa y no tiene demostración estadística, aunque generalmente se satisface por conveniencia, como es el caso particular de los ejemplos aquí presentados.
- En particular, cuando DEA se usa como una herramienta de evaluación comparativa (*benchmarking tool*), la técnica se enfoca en el desempeño individual de la DMU de interés y no necesariamente se debe satisfacer la regla de oro, y en este caso los resultados de DEA siguen siendo valiosos [Cook et al.2014].
- Antes de aplicar DEA debe estar claro qué se pretende alcanzar con el análisis. Si el objetivo es identificar las unidades que están sub-utilizando los recursos, entonces la reducción de las entradas será el foco del ejercicio y en tal caso los modelos DEA

orientados a entradas serán más apropiados. También es posible que el objetivo esté encaminado a incrementar los resultados de las unidades a partir de los recursos ya disponibles, lo cual hace pensar que los modelos DEA orientados a salidas son los ideales. Para el caso en el cual ambos objetivos son prioritarios el modelo DEA aditivo puede ser una solución correcta. En tal sentido, en nuestros ejemplos cualquiera de las tres visiones puede ser argumentada como adecuada. Sin embargo, a partir de los resultados de metodología propuesta, la empresa ineficiente logra asemejarse al conjunto referente de empresas eficientes con menos esfuerzo cuando el modelo de eficiencia evaluado fue el CCR-IO para C22 y CCR-OO para el C14, pero este resultado no es una condición general, y es posible que el modelo de eficiencia que mejor se adecua varíe según la empresa ineficiente de prueba, por tanto, la selección del tipo de modelo de eficiencia debe soportarse en la capacidad de intervención que la administración tenga sobre las variables relevantes y las políticas de gobierno previamente definidas.

- La búsqueda de variables relevantes a través de CCA redujo, no solo la carga computacional de los cálculos posteriores en el proceso de optimización de proyección sobre la envolvente convexa, sino que esencialmente permite disminuir la cantidad de intervenciones que debería realizar la gerencia, una vez sean determinados los resultados de la evaluación comparativa, en pro de transformar la DMU ineficiente en eficiente.
- En CCA es fundamental verificar la significancia estadística del análisis, porque esta prueba demuestra que los patrones de relación hallados entre las variables de entrada y la eficiencia no es cuestión del azar, confirmando que las variables relevantes deben ser el foco de atención para la administración de las DMUs.
- Es muy importante refinar las DMUs eficientes, seleccionando un subconjunto de las mismas, con el fin de garantizar que la envolvente convexa, que se construye posteriormente a partir de estas observaciones, no encierre DMUs ineficientes. Para el caso en que las funciones de densidad de probabilidad, para las DMUs eficientes e ineficientes, estén claramente separadas no se hace perentorio realizar este proceso.
- La gran ventaja de la estrategia de evaluación comparativa propuesta en este documento se basa en dos elementos: 1) la DMU ineficiente tendrá que hacer la menor cantidad de cambios posibles en sus variables relevantes de entrada para asemejarse al conjunto de empresas eficientes, 2) las variables de entrada pueden ser tanto decrementadas (caso habitual del *Benchmarking DEA*) como incrementadas para lograr un comportamiento semejante al de las empresas evaluadas como eficientes de referencia.
- En caso que la cantidad seleccionada de DMUs eficientes de referencia sea muy baja, es importante reducir o incluso eliminar el parámetro de umbral, y así permitir que más DMUs eficientes sean escogidas para el grupo de referencia favoreciendo la capacidad de generalización de los resultados.

Capítulo 8

Conclusiones

En esta investigación se propuso una metodología de análisis de relevancia de variables financieras por medio de la combinación de técnicas de análisis de frontera eficiente con técnicas de aprendizaje de máquina. Particularmente, son tres las contribuciones principales que contiene la propuesta metodológica.

En primer lugar, la metodología permite establecer cuáles de las variables de entrada son relevantes, es decir, determinar las variables que poseen mayor nivel de relación con la eficiencia de las DMUs, tal resultado se obtiene al aplicar el análisis de correlación canónica (CCA) entre las variables de entrada y la eficiencia calculada para el conjunto de DMUs que se analiza. Adicionalmente, una ventaja del análisis de correlación canónica es la posibilidad de calcular el nivel de significancia estadística, confirmando si los patrones de relación entre las variables relevantes y la eficiencia se presentan por cuestiones del azar o no. Como resultado, el número de dimensiones (variables) de análisis puede ser disminuido y así mismo pasa con los costos computacionales de etapas posteriores, pero más que esto, la selección de variables relevantes reduce la cantidad de intervenciones que deben realizar los cuerpos administrativos ayudando a focalizar su trabajo en una menor cantidad de elementos.

En segunda instancia, se planteó un método para refinar el conjunto de DMUs eficientes que sirven como referente para las DMUs ineficientes. Tal depuración es importante en tanto que permite eliminar del conjunto de datos las DMUs que aunque categorizadas como eficientes tienen un compartimiento muy similar a las DMUs ineficientes, es decir, se eliminan las DMUs que pueden generar confusión en la clasificación de eficiencia.

Y como tercera contribución de este trabajo, y tal vez la más importante, se propuso un novedoso método de evaluación comparativa, denominado proyección de distancia mínima (MDM), el cual posibilita calcular los valores de cambio para cada una de las variables (relevantes) de una DMU ineficiente, de manera que ésta pueda asemejarse a un conglomerado de empresas eficientes, brindando de esta manera un sistema de aprendizaje de máquina que determina la mejores posibilidades para la DMU ineficiente. La proyección de distancia mínima se fundamenta en minimizar la distancia entre una DMU ineficiente (en evaluación) y la envolvente convexa construida a partir de las DMUs eficientes que se empleen como referencia. La evaluación comparativa por proyección de distancia mínima tiene dos grandes ventajas: 1) la DMU ineficiente tendrá que hacer la menor cantidad de cambios posibles en sus variables relevantes de entrada para asemejarse al conjunto

de empresas eficientes, y 2) las variables de entrada pueden ser tanto decrementadas como incrementadas para lograr un comportamiento semejante al de las DMUs eficientes de referencia. Este segundo punto es un factor fundamental de diferenciación entre nuestra metodología MDM y la evaluación comparativa de DEA (*benchmarking DEA*). En *benchmarking DEA* sólo es posible de decrementar las variables de entrada, pero nunca se puede incrementarlas. En mercados competitivos la hipótesis de incrementar variables para mejorar las condiciones de eficiencia es común y perfectamente válida.

Adicionalmente, es importante mencionar que la metodología general propuesta no solo es útil para la evaluación y comparación de variables financieras, sino que puede ser empleada en otras áreas de interés, estableciendo adecuadamente las variables de entrada y salida que evalúen la eficiencia de una DMU.

Capítulo 9

Trabajo futuro

Siguiendo la línea de investigación que se ha descrito en esta tesis, surgen cuatro proyectos principales que pueden fortalecer y extender el alcance de la metodología propuesta.

1. Dado que la proyección de distancia mínima (MDM) no tiene en cuenta las variables de salida, bajo el supuesto de que esta información está codificada en las relaciones y etiquetas de eficiencia, es necesario durante la evaluación comparativa estimar un valor objetivo (valor de referencia) para las variables de salida.
2. La selección del parámetro umbral para refinar la selección de DMUs eficientes se basa en ensayo y error, y se soporta en la experticia del usuario evaluando las gráficas de las funciones de densidad de probabilidad de los conjuntos de DMUs eficientes e ineficientes. En tal sentido es necesario plantear una estrategia para optimizar el parámetro umbral o plantear una alternativa del proceso de selección de DMUs eficientes que garantice que ninguna DMU ineficiente quede encerrada por la envolvente convexa.
3. Evaluar la posibilidad de agregar un conjunto adicional de restricciones en el proceso de optimización MDM, que tenga en cuenta los casos, en que por condiciones de lógicas, las variables estén obligadas a solo incrementar o decrementar.
4. Aplicar la metodología propuesta en este trabajo sobre otro tipo de datos y verificar la estabilidad y congruencia de los resultados, idealmente con el propósito de demostrar que los valores objetivo de las variables resultan de utilidad para los administradores.

Bibliografía

- [Assaf and Josiassen2016] Assaf, A. G. and Josiassen, A. (2016). Frontier analysis: A state-of-the-art review and meta-analysis. *Journal of Travel Research*, 55(5):612–627.
- [Banker et al.1984] Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9):1078–1092.
- [Biddle et al.1997] Biddle, G. C., Bowen, R. M., and Wallace, J. S. (1997). Does eva® beat earnings? evidence on associations with stock returns and firm values. *Journal of accounting and economics*, 24(3):301–336.
- [Charnes et al.1985] Charnes, A., Cooper, W. W., Golany, B., Seiford, L., and Stutz, J. (1985). Foundations of data envelopment analysis for pareto-koopmans efficient empirical production functions. *Journal of econometrics*, 30(1-2):91–107.
- [Charnes et al.1978] Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.
- [Cook et al.2014] Cook, W. D., Tone, K., and Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, 44:1–4.
- [Cooper et al.2000] Cooper, W. W., Seiford, L. M., and Tone, K. (2000). Data envelopment analysis. *Handbook on Data Envelopment Analysis, 1st ed.*; Cooper, WW, Seiford, LM, Zhu, J., Eds, pages 1–40.
- [de Wet2005] de Wet, J. H. (2005). Eva versus traditional accounting measures of performance as drivers of shareholder value—a comparative analysis. *Meditari Accountancy Research*, 13(2):1–16.
- [Dinero2017] Dinero (2017). Los retos que enfrentan las mipymes en colombia. <https://www.dinero.com/edicion-impresa/pymes/articulo/los-retos-que-enfrentan-las-mipymes-en-colombia/241586>.
- [Espinosa et al.2015] Espinosa, F. R., Molina, Z. A. M., and Vera-Colina, M. A. (2015). Fracaso empresarial de las pequeñas y medianas empresas (pymes) en colombia. *Suma de negocios*, 6(13):29–41.

- [Farrell1957] Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, 120(3):253–281.
- [Feltham et al.2004] Feltham, G. D., Issac, G. E., Mbagwu, C., and Vaidyanathan, G. (2004). Perhaps eva does beat earnings—revisiting previous evidence. *Journal of Applied Corporate Finance*, 16(1):83–88.
- [Fenyves et al.2015] Fenyves, V., Tarnóczy, T., and Zsidó, K. (2015). Financial performance evaluation of agricultural enterprises with dea method. *Procedia Economics and Finance*, 32:423–431.
- [Feroz et al.2003] Feroz, E. H., Kim, S., and Raab, R. L. (2003). Financial statement analysis: A data envelopment analysis approach. *Journal of the Operational Research Society*, 54(1):48–58.
- [Golany and Roll1989] Golany, B. and Roll, Y. (1989). An application procedure for dea. *Omega*, 17(3):237–250.
- [Hui et al.2015] Hui, E. C., Gao, Y. O., and Chan, K. K. K. (2015). Does eva truly reflect the performance of property companies in china? *International Journal of Strategic Property Management*, 19(3):260–270.
- [Jakub et al.2015] Jakub, S., Viera, B., and Eva, K. (2015). Economic value added as a measurement tool of financial performance. *Procedia Economics and Finance*, 26:484–489.
- [Ji and Lee2010] Ji, Y.-b. and Lee, C. (2010). Data envelopment analysis. *The Stata Journal*, 10(2):267–280.
- [Khalil et al.2015] Khalil, S., Mehmood, B., and Nisar, A. (2015). Cost efficiency of pakistani banking sector: a stochastic frontier analysis. *The Journal of Commerce*, 7(3):110.
- [Lee and Kim2009] Lee, S. and Kim, W. G. (2009). Eva, refined eva, mva, or traditional performance measures for the hospitality industry? *International Journal of Hospitality Management*, 28(3):439–445.
- [Madsen et al.2004] Madsen, K., Nielsen, H. B., and Tingleff, O. (2004). *Optimization with Constraints*. Informatics and Mathematical Modelling – Technical University of Denmark, 2nd edition.
- [Méndez2007] Méndez, C. A. B. (2007). Ebitda, ¿es un indicador financiero contable de agregación de valor? *Capic Review*, (5):4.
- [Nataraja and Johnson2011] Nataraja, N. R. and Johnson, A. L. (2011). Guidelines for using variable selection techniques in data envelopment analysis. *European Journal of Operational Research*, 215(3):662–669.

- [O’Hanlon and Peasnell1998] O’Hanlon, J. and Peasnell, K. (1998). Wall street’s contribution to management accounting: the stern stewart eva® financial management system. *Management Accounting Research*, 9(4):421–444.
- [Pastor and Ruiz2007] Pastor, J. T. and Ruiz, J. L. (2007). Variables with negative values in dea. In *Modeling data irregularities and structural complexities in data envelopment analysis*, pages 63–84. Springer.
- [Ray2004] Ray, S. C. (2004). *Data envelopment analysis: theory and techniques for economics and operations research*. Cambridge university press.
- [Saha et al.2016] Saha, A., Ahmad, N. H., and Yeok, S. G. (2016). Evaluation of performance of malaysian banks in risk adjusted return on capital (raroc) and economic value added (eva) framework. *Asian Academy of Management Journal of Accounting & Finance*, 12(1).
- [Serrano and Blasco2000] Serrano, V. C. and Blasco, O. M. B. (2000). *Evaluación de la eficiencia mediante el Análisis Envolvente de Datos: Introducción a los modelos básicos*. B-EUMED.
- [Sexton et al.1986] Sexton, T. R., Silkman, R. H., and Hogan, A. J. (1986). Data envelopment analysis: Critique and extensions. *New Directions for Program Evaluation*, 1986(32):73–105.
- [Shi2012] Shi, Y. (2012). Eva and dea, which is better in reflecting the capital efficiency? In *Knowledge Discovery and Data Mining*, pages 15–23. Springer.
- [Shil2009] Shil, N. C. (2009). Performance measures: An application of economic value added. *International Journal of Business and Management*, 4(3):169–177.
- [Uurtio et al.2018] Uurtio, V., Monteiro, J. M., Kandola, J., Shawe-Taylor, J., Fernandez-Reyes, D., and Rousu, J. (2018). A tutorial on canonical correlation methods. *ACM Computing Surveys (CSUR)*, 50(6):95.
- [Wagner and Shimshak2007] Wagner, J. M. and Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European journal of operational research*, 180(1):57–67.